

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2000, 09:20:56 ; Search time 50.85 Seconds
(without alignments)
624.028 Million cell updates/sec

Title: US-09-428-122-2

Perfect score: 4774

Sequence: 1 MKSFFPKFVFSTFAIFPLSM.....FELRGSSRYNVNLDIAKYOF 928

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_36:*

1: /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*

2: /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*

3: /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*

4: /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*

5: /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*

6: /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*

7: /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*

8: /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*

9: /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*

10: /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*

11: /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*

12: /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*

13: /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*

14: /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*

15: /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*

16: /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*

17: /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*

18: /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*

19: /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*

20: /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

21: /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4774	100.0	928	21	Chlamydia pneumoniae
2	4760	99.7	928	20	Chlamydia pneumoniae
3	2802.5	58.7	597	20	Chlamydia pneumoniae
4	2058	43.1	928	20	Chlamydia pneumoniae
5	2006	42.0	918	21	Chlamydia pneumoniae
6	2001	41.9	918	20	Chlamydia pneumoniae
7	1993	41.7	914	20	Chlamydia pneumoniae
8	1986	41.6	928	20	Chlamydia pneumoniae
9	1940	40.6	930	20	Chlamydia pneumoniae
10	1927	40.4	930	20	Chlamydia pneumoniae
11	1917.5	40.2	927	20	Chlamydia pneumoniae
12	1856	38.9	949	20	Chlamydia pneumoniae

13	1855	38.9	928	20	Chlamydia pneumoniae
14	1832	38.4	945	21	Chlamydia pneumoniae
15	1811	37.9	945	20	Chlamydia pneumoniae
16	1572.5	32.9	643	20	Chlamydia pneumoniae
17	1564	32.8	841	20	Chlamydia pneumoniae
18	1430.5	30.0	922	20	Chlamydia pneumoniae
19	1417.5	29.7	922	20	Chlamydia pneumoniae
20	1377.5	28.9	973	21	Chlamydia pneumoniae
21	1259	26.4	671	20	Chlamydia pneumoniae
22	1214	25.4	220	20	Chlamydia pneumoniae
23	1203.5	25.2	507	20	Chlamydia pneumoniae
24	1130	23.7	1132	20	Chlamydia pneumoniae
25	1128	23.6	1012	20	Chlamydia pneumoniae
26	1110.5	23.3	1013	20	Chlamydia pneumoniae
27	1098.5	23.0	1013	20	Chlamydia pneumoniae
28	1015.5	21.3	450	20	Chlamydia pneumoniae
29	918	19.2	610	20	Chlamydia pneumoniae
30	888	18.6	530	20	Chlamydia pneumoniae
31	834.5	17.5	483	20	Chlamydia pneumoniae
32	821	17.2	494	20	Chlamydia pneumoniae
33	792	16.6	427	20	Chlamydia pneumoniae
34	768	16.1	177	20	Chlamydia pneumoniae
35	708	14.8	294	20	Chlamydia pneumoniae
36	680.5	14.3	1617	20	Chlamydia pneumoniae
37	667	14.0	947	21	Chlamydia pneumoniae
38	667	14.0	1146	20	Chlamydia pneumoniae
39	646	13.5	969	20	Chlamydia pneumoniae
40	635.5	13.3	400	20	Chlamydia pneumoniae
41	635	13.3	279	20	Chlamydia pneumoniae
42	603.5	12.6	940	20	Chlamydia pneumoniae
43	556	11.6	1194	20	Chlamydia pneumoniae
44	551	11.5	924	20	Chlamydia pneumoniae
45	546	11.4	989	20	Chlamydia pneumoniae

ALIGNMENTS

RESULT 1

ID Y94327 Standard: Protein: 928 AA.

AC Y94327;

DT 11-AUG-2000 (first entry)

DE Chlamydia pneumoniae 98kd putative outer membrane protein.

KW Chlamydia: antigen; vaccine; infection; outer membrane protein.

OS Chlamydia pneumoniae.

PN WO200026237-A2.

PD 11-MAY-2000.

PF 29-OCT-1999; 99MO-GB03579.

PR 29-OCT-1998; 98US-0106070.

PR 01-MAR-1999; 99US-0122066.

PR 27-OCT-1999; 99US-0428122.

(CONN-) CONNAUGHT LAB LTD.

PI Murdin AD, Oomen RP, Dunn PL;

DR WPI; 2000-365569/31.

DR N-PSDB; A27021.

PT Novel Chlamydia 98 kDa putative outer membrane protein antigen, used for vaccination and protection against Chlamydia infection

XX Claim 6; Fig 1; 93pp; English.

XX The present sequence is the 98kDa putative outer membrane protein from
 CC Chlamydia pneumoniae. The genomic sequence was amplified using two
 CC PCR primers. The 5' primer contains a NotI restriction site, a ribosome
 CC binding site, an initiation codon and a sequence close to the 5' end of
 CC the 98kDa putative outer membrane protein coding sequence. The 3' primer
 CC contains the sequence encoding the C-terminal sequence of the putative
 CC outer membrane protein and a BamHI restriction site. The stop codon was
 CC excluded and an additional nucleotide was inserted to obtain an in-frame
 CC C-terminal fusion with the Histidine tag. The PCR product was cloned
 CC into a eukaryotic expression vector (pCA-Myc-His) by restricting both
 CC the vector and the PCR product with NotI and BamHI and performing a
 CC ligation reaction. This expression vector was injected intramuscularly
 CC and intranasally into mice, which were subsequently inoculated with
 CC Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice
 CC were lower than those of the controls. Thus the 98kDa putative outer
 CC membrane protein can be used as a vaccine to provide protection against
 CC Chlamydia infections, especially Chlamydia pneumoniae infections.
 CC The present polypeptide may also be administered orally to treat
 CC Chlamydia infection.

SO Sequence 928 AA:

Query Match 100.0%; Score 4774; DB 21; Length 928;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKSFPKVFSTFAIFPLSMATETVLDSASFEDNKNNGFVSRESDAGTYLEKGV 60
 Db 1 mkssfpkfvfstfaifplsmatetvldsasfdgknngfvsresgedagtylfxgv 60
 OY 61 TLENIPGTGTAITKSCENNTKGDITFTGNNSLLFOTVDAGTAGAIVNSVVDKSTFI 120
 Db 61 tlenipgvtgtaitscfnntkgydlftgnngsllftqvvdagtagaivnsvvdksctfi 120
 OY 121 GFSSLSFASPSGSSITTGKGVSCSTGSLSTKKNVSLFSKNEFSTDGCAITAKTSLTG 180
 Db 121 gfsslsfaspssitgkgvscstgslstkknvslfsknfstdgcaitaktsltlg 180
 OY 121 gfsslsfaspssitgkgvscstgslstkknvslfsknfstdgcaitaktsltlg 180
 Db 121 gfsslsfaspssitgkgvscstgslstkknvslfsknfstdgcaitaktsltlg 180
 OY 181 TTMSALFSENTSSKKGAIQTSDALITTTGNOGEVSFSDNTSSDGAALFEASTISNNA 240
 Db 181 ttmsalfseentsskkgaitqtsdalittgnogevssfdntssdgaalfeastisnna 240
 OY 181 ttmsalfseentsskkgaitqtsdalittgnogevssfdntssdgaalfeastisnna 240
 Db 181 ttmsalfseentsskkgaitqtsdalittgnogevssfdntssdgaalfeastisnna 240
 OY 241 KVSFTDNKVTGASSSTTGDMGCAICAYKSTDPKVTLTGOMLFESNNTTGAIAIV 300
 Db 241 kvsftdnkvtgasssttgdmgcaicaykstdpkvtltgqomlfesnttgaiaiv 300
 OY 241 kvsftdnkvtgasssttgdmgcaicaykstdpkvtltgqomlfesnttgaiaiv 300
 Db 241 kvsftdnkvtgasssttgdmgcaicaykstdpkvtltgqomlfesnttgaiaiv 300
 OY 301 KKLIALSGILTLFSRNSVNGGTAPKGAIAIEDSGELSLADSGDIYFLGNTVTSTPGT 360
 Db 301 kklialsgiltlfsrnsvnggtapkgaiaiedsgelsladsdgiyflgntvtstpgt 360
 OY 301 kklialsgiltlfsrnsvnggtapkgaiaiedsgelsladsdgiyflgntvtstpgt 360
 Db 301 kklialsgiltlfsrnsvnggtapkgaiaiedsgelsladsdgiyflgntvtstpgt 360
 OY 361 NRSIDIGTSKMTALRSAGRAIFYPPIITGSSITVTDLKVNTPADALQYTGMI 420
 Db 361 nrsidigtakmtalrsagraifypipitgssitvtlknvntpadalqytgmi 420
 OY 361 nrsidigtakmtalrsagraifypipitgssitvtlknvntpadalqytgmi 420
 Db 361 nrsidigtakmtalrsagraifypipitgssitvtlknvntpadalqytgmi 420
 OY 421 FTGKELSTFEADSKNLTSLKLOPVTLSGGLSLKHGVTLQTOAFTQOADSRLMDYGT 480
 Db 421 ftgkelsftheadsknltslklopvtlsgglskhgvtlqtoaftqoadsrlmdygt 480
 OY 421 ftgkelsftheadsknltslklopvtlsgglskhgvtlqtoaftqoadsrlmdygt 480
 Db 421 ftgkelsftheadsknltslklopvtlsgglskhgvtlqtoaftqoadsrlmdygt 480
 OY 481 LEPADISTNNLVINISSIDGAKKAIETKATSKNLTLSGTTLLDPTGYENHSLNP 540
 Db 481 lepadistnnlviniissidgakkaietkatsknltlsgttllldptgyenhslnp 540
 OY 481 lepadistnnlviniissidgakkaietkatsknltlsgttllldptgyenhslnp 540
 Db 481 lepadistnnlviniissidgakkaietkatsknltlsgttllldptgyenhslnp 540
 OY 541 QSDYLIELKASGVTVSTAVTPPPIMGEKRYGOGTWGIVNGTASTTATNMKTGYI 600
 Db 541 qsdylieleksgvtvstavtpppimgkrygogtwgivngtasttatnmktyi 600
 OY 541 qsdylieleksgvtvstavtpppimgkrygogtwgivngtasttatnmktyi 600
 Db 541 qsdylieleksgvtvstavtpppimgkrygogtwgivngtasttatnmktyi 600
 OY 601 PNPERRGSIVPNSLWNAFIDISLHYLMETANEGLOGDAFWACAGISNFFHKDSTKTRG 660
 Db 601 pnperrgsivpnslnwafidislyhymetaneglogdafwacagisnffhkdstktrg 660
 OY 601 pnperrgsivpnslnwafidislyhymetaneglogdafwacagisnffhkdstktrg 660
 Db 601 pnperrgsivpnslnwafidislyhymetaneglogdafwacagisnffhkdstktrg 660
 OY 661 FRHLISGVIYIGNLHTCSDKILSAFCQLFGRDRDYFVAKNOGYVGGTLYYOHNETYIS 720

Db 661 frhlisgviyignlhtcsdkilsaafcglyfgrdryfvakngvtlyvgglyqhnetyis 720
 OY 721 LPCLRPSCLSYVTEPELVFSGNLSTHNDNLKRYTTPYVKSMGNDSPALFEGGR 780
 Db 721 lpclrpscslsyvtepelevfsgnlsythndnlkryttypvksmgndspalafeggr 780
 OY 781 APICIDSEALFEQYMPMKIQFYVAHOGFKEQTEAREFGSSRLVNLALPIGIRDKES 840
 Db 781 apicidsealeqympmkikfyvahogfkegtearefgssrlvnlalpigirdkes 840
 OY 841 DCQDATYNLGTVVDLVRSNPDCTTLRLISGDSMKRFGNRLAROLVIRAGNHFCENSN 900
 Db 841 dcqdatyntlgtvvdlvrsnppdcttlrlisgdsdkrlfgnrlarolviraghnfcnsn 900
 OY 901 FEARSQFEELRGSSRNPNVDLGAQYOF 928
 Db 901 fearsqfeelrgssrnpnvdlgakyof 928

RESULT 2

W88421
 ID W88421 standard; Protein; 928 AA.

AC W88421;

DI 26-APR-1999 (first entry)

DE Chlamydia pneumoniae surface exposed protein Omp8.

KW Omp8; outer membrane protein 8; surface exposed protein; antigen;
 infection; diagnosis; vaccine; atherosclerosis; asthma.

OS Chlamydia pneumoniae.

PN W09858953-A2.

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK00266.

PR 23-JUN-1997; 97DK-0000744.

PA (BIRK/) BIRKELUND S.
 (CHRI/) CHRISTIANSEN G.

PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 Mykland P;

DR WPI; 1999-105610/09.
 N-PSDB; X06820.

Species-specific test for identifying mammals infected with
 Chlamydia pneumoniae - comprises detecting antibodies specific for
 outer membrane proteins of C. pneumoniae or nucleic acids encoding
 these proteins

Claim 7; Page 53-55; 115pp; English.

This polypeptide comprises the novel 90.0 kDa surface exposed
 protein Omp8 of the human respiratory pathogen Chlamydia
 pneumoniae. Its amino acid sequence was deduced from DNA (see
 X06820) isolated from a C. pneumoniae expression library. The
 invention provides 12 novel surface exposed proteins; Omp4-Omp15
 (see W88417-28), and nucleic acid sequences encoding them (see
 X06816-27). A new species specific test is claimed that is used
 to identify mammals (including humans) infected with Chlamydia
 pneumoniae. The test comprises detecting antibodies specific for
 Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
 membrane proteins, especially by PCR. The proteins are also used
 in the diagnosis of C. pneumoniae infection in mammals. The
 nucleic acids and proteins can also be used in the immunization of
 mammals, the nucleic acids being particularly useful as DNA

CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
XX
XX
50 Sequence 928 AA;

Query Match 99.7%; Score 4760; DB 20; Length 928;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 925; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY 1 MKSSPKVEVFSEFAFPMTATETVLDSASFQDNKNGNFSVRESQDAGTTFYFKGNV 60
DB 1 mksspkvtfstfaifpmslatevldsasfdgnknfnsvresqdaqctylfkgnv 60
OY 61 TLENIPGTGTAITKSCFNNTKGLFTTNGNSLFPQTVAGTACAAVNSVVDKSTTFI 120
DB 61 tlenipgtgtaikscfnntkglfttngnslfpqtvagtacaaavnsvvdksttfi 120
OY 121 GFSSLSFIASPGSSITTTGKAVSCSTGSLTKKNVSLFSKNFTDNGATITAKLSLTG 180
DB 121 gfsslsfiaspssitttgkavscstgsltkknvslfsknftdngatitaklsltg 180
OY 181 TTMALFSENTSSKKGAIoTSDALITGNOGEVFSFNTSSDSCALETFEASVTISNNA 240
DB 181 ttmalfseentsskkgaiotSDALITGNOGEVFSFNTSSDSCALETFEASVTISNNA 240
OY 241 KVSFIDNKVTCASSSTTGDMSGAICAYKSTDTKVTLLTGNOMLFSNNTSTTAGAIYV 300
DB 241 kvsfidnkvtcasssttgdmsgaicaykSTDTKVTLLTGNOMLFSNNTSTTAGAIYV 300
OY 301 KKLEIASGGLTFSRNSVNGTAPKGAIAIETDSELSLSDGDIVFLGNVTSTTPTGT 360
DB 301 kkleiasggltfSRNSVNGTAPKGAIAIETDSELSLSDGDIVFLGNVTSTTPTGT 360
OY 361 NRSIDLTSAKMTALRSAGRAIYFDPTTGSSTVTDLKVNERTPADSALQVTGNII 420
DB 361 nrsidltSAKMTALRSAGRAIYFDPTTGSSTVTDLKVNERTPADSALQVTGNII 420
OY 421 FTGEKLESTEAADSKNLTSLQPTLSGGLTSLKHGVTLOTQAFQADSRLENDVTT 480
DB 421 ftgekleSTEAADSKNLTSLQPTLSGGLTSLKHGVTLOTQAFQADSRLENDVTT 480
OY 481 LEPADTSTNNLVINISSIDGAKKAKIEFKATSKNLTSGTTLDPGTGFEVENSINP 540
DB 481 lepadtSTNNLVINISSIDGAKKAKIEFKATSKNLTSGTTLDPGTGFEVENSINP 540
OY 541 QSYDILELKASGVTSTAVTPDPIMGEKFXHYGQGTWGPVWGTGASTTATFNMTKTYI 600
DB 541 qsydileLKASGVTSTAVTPDPIMGEKFXHYGQGTWGPVWGTGASTTATFNMTKTYI 600
OY 601 PNPERIGSLVNSLIMNAIDISSLHYMETANEGIQGRATWAGLSEPFKDKSTKTRG 660
DB 601 pnperigSLVNSLIMNAIDISSLHYMETANEGIQGRATWAGLSEPFKDKSTKTRG 660
OY 661 FRHLSCGVIVIGNLTCTCDKILISAACOLFGDRDRDYFAKNOGYVGGTLYVONHETIIS 720
DB 661 frhlscgvivignltctcdkILISAACOLFGDRDRDYFAKNOGYVGGTLYVONHETIIS 720
OY 720 661 frhlsgvivignltctcdkILISAACOLFGDRDRDYFAKNOGYVGGTLYVONHETIIS 720
DB 720 661 frhlsgvivignltctcdkILISAACOLFGDRDRDYFAKNOGYVGGTLYVONHETIIS 720
OY 721 LPCKLRPCSLSYVTEIPVLFSGNLSTYHTDNDLKTCTTYTPYKSGNDSFALFEGGR 780
DB 721 lpcklrpcslsyvteipVLFSGNLSTYHTDNDLKTCTTYTPYKSGNDSFALFEGGR 780
OY 781 APICIDSEALFEQYWPBKLOFYVAHOGCFKEOGTEAREPSSRVLNALITGRDKES 840
DB 781 apicidseALFEQYWPBKLOFYVAHOGCFKEOGTEAREPSSRVLNALITGRDKES 840
OY 841 DCODATYNTLTGYVDLVRSPDCTTTLRISGDSKWTGTLARQALVLRAGNHCFEFSN 900
DB 841 dcoDATYNTLTGYVDLVRSPDCTTTLRISGDSKWTGTLARQALVLRAGNHCFEFSN 900
OY 901 FEAFSQSFELRGSSRNINVDLGAQYF 928
DB 901 feafsqsfELRGSSRNINVDLGAQYF 928

```

DB 901 feafsqsfelrgssrninvdlgakqf 928

RESULT 3
ID Y34611
Y34611 standard; Protein: 597 AA.

XX AC Y34611;
XX DT 13-SEP-1999 (first entry)
XX

DE Chlamydia pneumoniae transmembrane protein sequence.

XX Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis;
XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN W09927105-A2.

PN 03-JUN-1999.

PE 20-NOV-1998; 98WC-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

PA (GSET) GENSET.

PI Griffais R;

DR WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Page 632-633; Disclosure; 1912pp; English.

CC Y34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see X91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.

SO Sequence 597 AA;

Query Match 58.7%; Score 2802.5; DB 20; Length 597;
Best Local Similarity 93.8%; Pred. No. 6,5e-178;
Matches 560; Conservative 5; Mismatches 21; Indels 11; Gaps 3;

```

OY 149 LSLTKNNVSLFSKRFSTNDNGATITAKTLSLTGTTMSALFSENNSSKKGAIOIOTSDALIT 208
DB 1 lsltknnvslfskrfstndngatitAKTLSLTGTTMSALFSENNSSKKGAIOIOTSDALIT 208
OY 209 GNOGEVFSFNTSSDSCALETFEASVTISNNAKVSFDNKVTCASSSTTGDMSGAICAY 268
DB 61 gnggevfsfntssdscALETFEASVTISNNAKVSFDNKVTCASSSTTGDMSGAICAY 268
OY 269 KTSTDITVTLTGNOMLFSNNTSTTAGAIYVKKLEIASGGLTFSRNSVNGTAPKGA 328
DB 121 ktstditvTLTGNOMLFSNNTSTTAGAIYVKKLEIASGGLTFSRNSVNGTAPKGA 328
OY 329 IATDSELSLSDSGDIVFLGNVTSTPTGTRSSIDLTSAKMTALRSAGRAIYFD 388
DB 181 iatdseLSLSDSGDIVFLGNVTSTPTGTRSSIDLTSAKMTALRSAGRAIYFD 388

```

QY	389	PLTGSSTVTMDVLKVNETFPADSAQYTGNIITFGEKISTEADSKNLTSLQPTTIS	448
Db	241	plltgssltvtvllkvnepadaslqyfgnllftgekiseeadsknltskllqptlls	300
QY	449	GGTSLKSGVTLQOTAFQTQADSRLEMOVGTLEPADSTNNINVINISSIDGAKKAKIE	508
Db	301	ggtlalkhgvltqlqafatqgadsrlendvgltlepadstnnlnlvlnissldgakkakie	360
QY	509	TRATSKNLTSLSTITLLDPTGTFYENHSLRNPQSDIIELKASGTVTSTAVTPDPIMGEK	568
Db	361	tratsknltlsqtlclltpqtlfyenhsllrnpqsydllelkasgtvtstavtpdpimgek	420
QY	569	FHYGVGTGPIVWGTGASTATFEMWTKGTGYPPIPERIGSLVPSLWNAFLDISSLHYLM	628
Db	421	fhyygvgtgwplvwgtgastatfemwtktylppperigslvpslwnafldisshylm	480
QY	629	ETANEGLOGDRAFWCAGLSNFFHKDSTKTRRGRFHLSGGVYIGGNLTCSDKLTLSAFCQ	688
Db	481	etaneaglgdratfwcaglsnffhkdstktrrgfhlsggyvlggnlhtcsdkllsaafcg	540
QY	689	LEGRPRDVFVAMKNGS-TYGGGILVYQNHMYIS-----LPCKLR-PCSLSTYP	734
Db	541	lgrprdyfvamkngyslrtnslpagnlylsslqtlatfvlvcysydsclfrx	597
RESULT 4			
ID	W88418	W88418 standard; Protein; 928 AA.	
XX	W88418:		
XX	26-APR-1999	(first entry)	
DE	Chlamydia pneumoniae surface exposed protein Omp5.		
KN	Omp5: outer membrane protein 5; surface exposed protein; antigen;		
KN	infection; diagnosis; vaccine; atherosclerosis; asthma.		
OS	Chlamydia pneumoniae.		
PN	MO9858953-A2.		
XX	30-DEC-1998.		
XX	19-JUN-1998;	98WO-DK00266.	
PR	23-JUN-1997;	97DK-0000744.	
XX	(BIRK/) BIRKELJUND S.		
PA	(CHR1/) CHRISTIANSEN G.		
PI	BirkeLund S, Christiansen G, Knudsen K, Madsen A;		
PI	Myind P;		
DR	WPI: 1999-105610/09.		
DR	N-PSDB: X06817.		
XX	Species-specific test for identifying mammals infected with		
PT	Chlamydia pneumoniae - comprises detecting antibodies specific for		
PT	outer membrane proteins of C. pneumoniae or nucleic acids encoding		
PT	these proteins		
CC	Claim 7; Page 43-45; 115pp; English.		
CC	This polypeptide comprises the novel 97.2 kDa surface exposed		
CC	protein Omp5 of the human respiratory pathogen Chlamydia		
CC	pneumoniae. Its amino acid sequence was deduced from DNA (see		
CC	X06817) isolated from a C. pneumoniae expression library. The		
CC	invention provides 12 novel surface exposed proteins, Omp4-Omp5		
CC	(see W88417-28), and nucleic acid sequences encoding them (see		
CC	X06816-27). A new species specific test is claimed that is used		
CC	to identify mammals (including humans) infected with Chlamydia		
CC	pneumoniae. The test comprises detecting antibodies specific for		

CC	OmpA-OmpL5 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used
CC	in the diagnosis of C. pneumoniae infection in mammals. The
CC	nucleic acids and proteins can also be used in the immunization of
CC	mammals, the nucleic acids being particularly useful as DNA
CC	vaccines for effecting in vivo expression of antigens. The
CC	vaccines may also prevent atherosclerosis and bronchial asthma,
CC	which are possibly associated with C. pneumoniae.
XX	
SQ	Sequence 928 AA:
	Query Match 43.1%; Score 2058; DB 206; Length 928;
	Best Local Similarity 46.7%; Pred. No. 2.8e-126;
	Matches 444; Conservative 142; Mismatches 320; Indels 44; Gaps 20;
QY	1 MKSSPFKEFV-STFAIF--PLSMATETVLDDSSAFDKNK-GNFVSREQEDAGTYIL 55
Db	1 mksqfwmvlslstlactscsfvataenaispsdsfgstntgcytknt--ttgidyt 58
QY	56 FKGNTYLENIEPQTGAITSCFNNTKGDLPFGNGSLFPQTVDAGTVAAVNSVYDK 115
Db	59 lvgdltlgnl-dsaalrtgcfdctdesisfigkyslsflnks-saeagal-vytlck 115
QY	116 STTFIIFSSLSTIASPGSSITT--CKGAVSCGTGSLSLTKNVLSFKNFSDNCAGTIA 173
Db	116 nslsgfssltllaapsrvlttspskgavkgv-gdltdmglltfkdqdeemngaist 174
QY	174 KTLSTLGTTMSALFSEBNTSS--KKGAIQISDALTTIGNOGVEFSNTSSDCAIIFT 230
Db	175 krlslkstgsisfignkxsatgkkggalcatgyvtldtnaptlfmsiaaagaans 234
QY	231 EASVISNNAKRSFDNKVTGASSSTGMSCGACAKTSRDTKVTLTGNMQLFSNNT 290
Db	235 tgnclitgnlstvfienstv----atag--nggal-----sgdaavltsngsvfvsgnq 283
QY	291 STTAGAGIVVKLELAS--GGTLFESRVNSVNGTAPKCGAIIAIEDSGELSLADSGDIVF 348
Db	284 avangagaiyakklltasgggggisfsmlyvggttagngaisllaagecslaeaqidltf 343
QY	349 LGNTYTSTTP--GTNSSIDIGSAMTMLRSAGAIFYEYPRTIGSSITYTDLVKIVET 407
Db	344 ngnaivaatcpgtkrnsidigstacilrlraasngsiftfyptiantaadcdlnluka 403
QY	408 PADSALOYTGNIIFIPEGEKISETEADSKNLTKLOPYTLGGCTLSKHGYTLOTQATO 467
Db	404 dagnstdysivsfegeklsedeakvadnltslkpqrtlagnhlvkrtyltdtkgftg 463
QY	466 QADSLEMDVGITLEPA-DSTINLVINISSIDGAKAKIETKATSNLLTSGTITLLD 526
Db	464 tagssvindhagltlkasteeyelltgls.lpvdsjgegkkvvlaasaasknvlaspllld 523
QY	527 PGTGYENHSHLSNPQSYDILELKAGTYTAIVDPDPMGEKFHHYGCTMGPIYW--- 582
Db	524 ngmaayenhdkigtgdfsvqlsalgtattctdpavrcvaparthygygtvgv-mtwaddt 582
QY	583 -GTGSTATAFERWKTGITIPNPERIGSLVPNSLMNAFLDISLHYLMETANEGIOGDRAF 641
Db	583 astprktlatlatwltgylipnperrgpiyvpalslwgfsfedigaigvleratslittcsdrif 642
QY	642 WAGAGSNFFEHDSSTRTRGRGFHLGSGYVIGGNLHMCHCSKLIISAFCOLFGDRDRIYAKN 701
Db	643 waagvaanfldkdkkyekrkyyhnksagyaiggaatceesnlstfcqlfgsdckflvaqn 702
QY	702 OGTYVGGTLYYQHNETYISLPCRLAPCSLSYV--TEIPVLFSGNLSYTHTDNDLKITKY 758
Db	703 htdtyagafiyqh---itecsfigiclldklpsswnshkprlvleqglayshvndkltkxy 758
QY	759 TTYPRVKSWMGNDALFEFGRAPICLDESKALFEQYMPFMKLQCVYAHOEFKROGFIAR 818
Db	759 tayevvksswnaaetnmmlqasshsypeylnctictyparyiklnltlylrtdgsfisekgtegr 818
QY	819 EFGSSRLVNALPIGIRPDKESCDODAYNNLTLYVDLVNSNDPCYTTLRISDSMKTFF 878


```
Db      819 sfdsnlfmislpryfkfscndfisyallfslsyvdlrlnobpkctcalvisgaswely 878
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      879 GTNLAROLVLRAGNHCFNSNEAFSOFSEFLRGSSRNYNVDLGAKYQF 928
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      879 annlarqalqvragshyafspmfefvlgqfvefyvgsrrlrynvdlgkfqf 928

RESULT  5
Y69369  Y69369 standard; Protein: 918 AA.
AC      Y69369;
XX      19-JUN-2000 (first entry)
XX      Amino acid sequence of the CPN100395 polypeptide.
DE      CPN100395; Chlamydia infection; immune response; vaccine.
XX      Chlamydia pneumoniae.
OS      WO200011183-A2.
XX      02-MAR-2000.
XX      18-AUG-1999: 99WC-IB01449.
XX      20-AUG-1998: 98US-0097187.
XX      20-AUG-1998: 98US-0097188.
XX      20-AUG-1998: 98US-0097189.
XX      20-AUG-1998: 98US-0097190.
XX      20-AUG-1998: 98US-0097195.
XX      20-AUG-1998: 98US-0097196.
XX      20-AUG-1998: 98US-0097197.
XX      20-AUG-1998: 98US-0097191.
XX      17-AUG-1999: 99US-0376770.
XX      (CONN-) CONNUGHT LAB LTD.
XX      Murdin AD, Oomen RP:
XX      WPI: 2000-224703/19.
XX      N-PSDB: Z61509.
XX      Novel antigens and corresponding DNA molecules that can be used to
PT      prevent, treat and diagnose disease caused by Chlamydia infection in
XX      mammals, especially humans -
XX      Claim 19; Fig 15-E; 201pp: English.
XX      Y69362-69 represent Chlamydia pneumoniae polypeptides. The polypeptides
CC      are present in the bacterial membrane structure, in the external
CC      vicinity of the membrane structure, in the inclusion membrane
CC      structure, in the external vicinity of the inclusion membrane structure,
CC      and in the cytoplasm of the infected cell. The polypeptides may be
CC      used to prevent, treat and detect the presence of Chlamydia infection
CC      and/or the presence of Chlamydia in a sample. The polypeptides may
CC      also be used to induce an immune response in a mammal. The vaccine
CC      vector comprising the polynucleotides is used to induce an immune
CC      response in a mammal. Antibodies directed against the polypeptides
CC      may also be used therapeutically to treat and/or prevent a Chlamydia
CC      infection.
XX      Sequence 918 AA:
SQ
```

```
Query Match      42.0%; Score 2006; DB 21; Length 918;
Best Local Similarity 45.6%; Pred. No. 7.7e-125;
Matches 432; Conservative 150; Mismatches 316; Indels 50; Gaps 20;
QY      1 MKSSFKFVSTFAIFPLSMI---ATETVLDSASFDGN-KKGNVSIVRSQEDA-GTTY 54
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db      1 mssfcllllssslapllmsvsadaatlfsrdsyngdtstfctfpaaksdaagty 60
QY      55 LFKGNVTLLENIPOTGTAIRKSCFNNTKGDLTFTGNGNSLLFQTVDAITVAGAANSSVVD 114
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 lldgdrvisq-agkqslstscfsmagnlftfingfshfchfnllsstvaygvnshtaas 119
QY      115 KSTTFGFSSLSLFIAPSGSITTTGKAVSCSTGSLSLTNVSLSLFSKNSTDNCAITRK 174
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      120 gltkisgfslrlmaapr---tlgkgaikltcg-lvfeisignldnneassenggaank 175
QY      175 TSLTGTMTSALFSENTSSKKGAISOTSDALITTGNGEVSFDNTSSDGAIFFEASV 234
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      176 tllstgstrfvalignssqggaiysagdsvisenagllsfmsnatsggaiasegnl 235
QY      235 TISNNAKVSFIDNKVTGASSSTTGDSGCAICAYK--TSTDTRKVTLTGNOMLFSNNTST 292
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      236 vlsnngniffdgkcat-----tngaidcnkaganpdrllttsnesihflnntag 286
QY      293 TNGAIIYVKLELASG-GLTFSRNVNGCTAPKGAIAIEDSGELSLSDSGDIYFVLCN 351
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      287 nsqgaalytkkllvlsgrgylfsmkaanat-pkgaialdsgeislsgadlignllfeqn 345
QY      352 TVTSTT--PGTNRSSIDLTSAKMTALRSAGRAIFYDPITTGSTTVTDVLKVNTP 408
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      346 t-sttgspsasvvrnaidlasnakflnlratrgnkvlfydplt---ssgatklslnked 401
QY      409 ADSALQYTGNIITFGKLSLTFEADSKNLTSKLQPVLTISGGLSLKSHVTLQTOAFYQ 468
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      402 agsgnlyegylvtsgeklseelkkpdklktftgvealaagalykddvrvvaantltcv 461
QY      469 ADSRLMDVQTTLE-PADSTINLVINISIDGAKAKIEKATSKNLTSGTITLDP 527
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      462 egskvmdggtlfeasaegvlnlgnlndldgntkalkataaskdvalsgplmva 521
QY      528 TGTTFEYHSLRNPQSDIILEKASCTVSTAVTPDIMEKHYGQGWGPIW---GT 584
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      522 qgnyehnlsgqvfprielsagctmtdldpdlrlctthhyqgnwn-lvwvddat 580
QY      585 GASTTATFMNTRKGYIPNERIGSLVPSNLSMNAFIDISSLHLMETRANGLOGDRAFWCA 644
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      581 aktncaltwtkgykpnpergplvpsnlwgsfydvrsigslnmdstsslsstcnlws 640
QY      645 GLSNFPHKDSIRKGRFRLSGGVYVGNLHTCSDKILSAFCOLFGRRDYPVAKNOST 704
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      641 gladfhedqkngqrsyzrsasgalyggfftasenfftfatcqlfygkdhlvakhnh 700
QY      705 VYGGTLYIQH---NETYISLPCKLRPCSLSYPTLIPVLFSGNLSYTHTDNDLTKITYTY 761
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      701 vygamsyrhlgesktlaktl-----lsgnsdsipfinafaygldnmtlkyty 752
QY      762 PTYKSGMNDSFALPFGGRAPICLD-ESALFEOYMPFMKLOFYVAHOEGFEGTAREAF 820
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      753 spyksgwgdafgieeggaipvaasrrsvdthpdlrlnemayahqndfkgfegzrf 812
QY      821 GSSRLVNLALPIGIREDKESDCODATYNTLTGYVLDVLRNSPDCTTTLRISGDSMKTFCT 880
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      813 gseidlflavpyikfektfd--kstydisiaypovlrindpgetctllmvsdswstcgt 870
QY      881 NLAROLVLRAGNHCFNSNEAFSOFSEFLRGSSRNYNVDLGAKYQF 928
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      871 slsrqallvraqnhaafasnfefsgfevelrgssrsyaldlgigtgf 918

RESULT  6
W88422  W88422 standard; Protein: 918 AA.
XX      W88422;
XX      26-APR-1999 (first entry)
XX      Chlamydia pneumoniae surface exposed protein omp9.
DE      XX
```

Om9: outer membrane protein 9; surface exposed protein; antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.
 Chlamydia pneumoniae.
 W0858953-A2.
 30-DEC-1998.
 19-JUN-1998; 98WO-DK00266.
 23-JUN-1997; 97DK-0000744.
 (BIRK/) BIRKELUND S.
 (CHR1/) CHRISTIANSEN G.
 Birkelund S, Christiansen G, Knudsen K, Madsen A;
 Mygdin P;
 WPI: 1999-105610/09.
 N-PSDB; X06821.
 Species-specific test for identifying mammals infected with Chlamydia pneumoniae - comprises detecting antibodies specific for outer membrane proteins of C. pneumoniae or nucleic acids encoding these proteins
 Claim 7; Page 56-58; 115pp; English.
 This polypeptide comprises the novel 96.7 kDa surface exposed protein Om9 of the human respiratory pathogen Chlamydia pneumoniae. Its amino acid sequence was deduced from DNA (see X06821) isolated from a C. pneumoniae expression library. The invention provides 12 novel surface exposed proteins, Om4-Omp15 (see W88417-28), and nucleic acid sequences encoding them (see X06816-27). A new species specific test is claimed that is used to identify mammals (including humans) infected with Chlamydia pneumoniae. The test comprises detecting antibodies specific for Om4-Omp15 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of C. pneumoniae infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with C. pneumoniae.

QY	288	NNISTAGCAIYVKKLELASG-GLTIPSRNVNGCTAPKGCALIIDSGEISLADSGDI	344
Db	282	mtaenssgaiaytkkllvssrggvlfsnmkaanat-pkgaiallidsgeistsadignl	340
QY	347	VELGTVTST---PGTRRSSIDIGTSAKMVALRSAGRAIYFDPDPTTGGSTVTIVLK	403
Db	341	lfeqnt-estsgpasvcrnmaidlasmakfmlnatgmkvifdydptl---ssgatkls	399
QY	404	VNEPDPADALOTGNIITFGKELSETEADSKNLTSKULOPVTLSGGTSLKHNVTLOTQ	463
Db	397	lnkadagsnutyeyjivfsgsklseeeelkxpdnlikstftgvevlaagalvldqdvtyvan	456
QY	464	AFTQADRLLEMDVGTGLE-PADSTININIVINISIDGAKAKIEFKANSKNTLSGTI	522
Db	457	cttgyegskvmdggtctleasaeyvlinglaidnsldgtlnkalikataaskdvalsgpl	516
QY	523	TLLDPTGFENHSLRNPQSDYLIELKASGTSTAVTPPEIMGEKPFHYGYOGTWGPVW	582
Db	517	mlvdagngnyehnlmsqgqvfprielsegumtcttdpdrlnltnthygqf-gllvw	575
QY	583	---GIGAATTATFNNTKTCGYIPNEERIGSLVPSNLMNAFIDISLHLMETANGLOGDR	639
Db	576	vddataclaktnatlwtcktykypnperqgplvpnslwsgivdvsisqslmdrstsslsst	635
QY	640	AFWCGELNPFHKNDSKTRRGFRHLISGGVYVIGNLHTCSDKILISAARCOLFGRRDVEFA	699
Db	636	nlwsgiaedfihbedkngnqrsyrssesaigaig9gfafsaenffirafctqdfygdkhalva	695
QY	700	KNOGTIVYGGTLTYOH---NETYISLPCKLRPCSLSYVPEIPIVLFSGNSLSTHTNDLKT	756
Db	696	knhtvaygamsyhlgsesklakl-----lsgnsdipfvinarfagphdnmmt	747
QY	757	KYTTYPTVYKSGWGNDSFALFEGGRAPICLD-ESALFEQYMPFMKLOQVYVAHQEGFKQGT	815
Db	748	kyltgyprpkysgwndaiglecg9aipvvaasgrtswdthcpflnlemlyahqndikengt	807
QY	816	EAREGSSRLVNLALPIGIRFDEKSDCODATYNTLTGVTVDLVRSNPDCTTTRISDSW	875
Db	808	egrtsgesedlnhlavpgrikrefsd--ketydaisiaypdrvinrdpgcttllmwsgdsw	865
QY	876	KTFGTNLARQALVLRAGNHFCFNSFEAFSOFSEFLGRSSRNMYVDLGARYOF	928
Db	866	stcgtsisrqaillvragnhhaafasnfevfsgfevelrgssrasyaidigrfgf	918
RESULT 7			
W88429	ID	W88429 standard; Protein; 914 AA.	
AC			
XX	W88429;		
XX			
DT	26-APR-1999	(first entry)	
XX			
DE	Chlamydia pneumoniae surface exposed protein Omp5.		
XX			
KW	Omp5; outer membrane protein 5; surface exposed protein; antigen;		
KW	infection; diagnosis; vaccine; atherosclerosis; asthma.		
XX			
OS	Chlamydia pneumoniae.		
XX			
OS	W09858953-A2.		
XX			
PD	30-DEC-1998.		
XX			
PF	19-JUN-1998;	98WO-DK00266.	
XX			
PR	23-JUN-1997;	97DK-0000744.	
XX			
PA	(BIRK/) BIRKJUND S.		
PA	(CHR1/) CHRISTIANSEN G.		
XX			
XX	Birkelund S, Christiansen G, Knudsen K, Madsen A;		
XX			

PI Mybind P:
 XX WPI; 1999-105610/09.
 DR N-PSDB; X06828.
 XX
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 XX
 PS Disclosure; Page 78-80; 115pp; English.
 XX
 CC This polypeptide is described as a subsequence of the claimed
 CC novel surface exposed protein omp5 (see W88418) of Chlamydia
 CC pneumoniae, a human respiratory pathogen. The invention provides
 CC a new species specific test for identifying mammals (including
 CC humans) infected with C. pneumoniae. The test comprises detecting
 CC antibodies specific for surface exposed proteins Omp4-Omp15 (see
 CC W88417-28) or detecting nucleic acid fragments encoding them (see
 CC X06816-27), especially by PCR. The proteins are also used in the
 CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids
 CC and proteins can also be used in the immunization of mammals, the
 CC nucleic acids being particularly useful as DNA vaccines for
 CC effecting in vivo expression of antigens. The vaccines may also
 CC prevent atherosclerosis and bronchial asthma, which are possibly
 CC associated with C. pneumoniae.
 XX
 XX Sequence 914 AA:
 SQ
 Query Match 41.7%; Score 1993; DB 20; Length 914;
 Best Local Similarity 46.2%; Pred. No. 5; 5e-124;
 Matches 432; Conservative 141; Mismatches 319; Indels 44; Gaps 20;
 QY 1 MKSSEPKFVF-STFAIF--PLSMIAETVLDSSASFDGKN-KNFSYRESQEDAGTTL 55
 DB 1 mksqfslvslstlaactscetvfaataenqpsdtsfntgtycpknt--ctgldyt 58
 QY 56 FKGNVTLENIPTGTATTKSCFNNTKGDITFTNGNSLIPQTVAGVYAGAAVSSVVDK 115
 DB 59 lqgdlitqlnlgdssaalqkqfcdtslstaqkqyslfnlks-aagagal-svtdtk 115
 QY 116 STPEIGFSSLSFIASPCSSITF--GKGVASCSTLSLTKNVSLFSGNFSTDNGAITA 173
 DB 116 nlsitgfsstlflaapsavltltpsgkqavkcg-qdlfdmngtllfkdyeeengalst 174
 QY 174 KTLULGTMTMSALFSEMTSS--KKGAIQTSDALTTGNOGEVFSQDNTSSDGAIFT 230
 DB 175 knlskntsgtsisfegnkksatgkkgalcgtvdltnlfnlaaagaagains 234
 QY 231 EASYTISNNKAVSFIDNKVYGASSSTTGDMGGAICAVKTSDDKVTLLTGOMLLEFNNT 290
 DB 235 tgnctlgnsislvsensvtc---atag--nggal-----sgdadvtlsqngsvltisng 283
 QY 291 STTAGAIIYKKLELAS--GGLTFSRNSVNGTAPKGAIAIEDSGELSLASDGDIVF 348
 DB 284 avanggaiaiyakklllaasggggaisfnmivgttagnggalslaagcscsaagdlft 343
 QY 349 LGNVITSTTP-GTNRSSIDGTSKMTALRSAGRAIYFDPTTGSSSTVTVDLKNET 407
 DB 344 ngnaivaltqtktrnsldigtstakltlnraishsiftdyplantaadstldlnlka 403
 QY 408 PADSALOYTGNIPTGEKLESTEADSKNLTKLQPVTSGGTSLKHGVTLOTQAFQ 467
 DB 404 dagnstlysgsivtsgeklseadeakvadnltstlkpvtltagtlvkrvltldtkftg 463
 QY 468 QADSRLMDVGTLEPA-DTSTINNLVINISSIDAKKAKIETAKSKNLTLSGTFITLD 526
 DB 464 tagsvvlmdagttlkasteetlglislpvdsldgkqvvaasaaknvalspillld 523
 QY 527 PTGFFYNNHSLRNQSYDILELKASGVTSRAVTPDPMGKRPYVGOGTMGPVW---- 582
 DB 524 ngnaayenhdlgkqcdsfvqlsdlgcatcttdpavpvtatphtlygygltwgt-mtwvddt 582

QY 583 -GTGASTTATFNWTKTGYPIDNPERISGLVPSNLMAFIDISSLHYLMEANEGLOGDRAP 641
 DB 583 aslptkxtlatwntngylpnpqrgplvpnsiwsfsgfdqaylversaltlcsdrfg 642
 QY 642 WCGGLSFFFKDSTKRRCGRHLSGCVIGGNLHTGSDKLTLSAFOQLFRDNDIYVAKN 701
 DB 643 waagvanfildkdkkgekrkyrksygaaygaagcseanlstafoqlfsgdcdflvakn 702
 QY 702 QGTWYGGTLYOHNETYISLPCRLPCSLSYVP---TEIPVLFSGNLSYRHTNDLTKTY 758
 DB 703 htdtyagafyiqh-----lccsgfigclldklpgswshkplvlegqayshvndlktky 758
 QY 759 TTYPTVKGSGNDSFALFEGGRAPICDSALREBOYWPENKLOFYVAHOGEFKEQTEAR 818
 DB 759 tayeveksgvgnafnmmlgaashsypeylhcdtyapyiklilyrdsfsekgtegr 818
 QY 819 EFGSSRLVNLALPIGIRFDESDCOADATYVLTGYVYDVRSNPDCTTTLRISGDSWTF 878
 DB 819 sfdsnlnfnslpkyvxfekfscndtsydltslypdlirnpkcttalvisgaewely 878
 QY 879 GTNLARQALVLRAGNHCFNSNEAFSQSFELRGS 914
 DB 879 annlarqalvragshyafspmfewlqgfvfevrgs 914
 RESULT 8
 W88423
 ID W88423 standard; Protein; 928 AA.
 XX
 AC W88423;
 XX
 DT 26-APR-1999 (first entry)
 XX
 DE Chlamydia pneumoniae surface exposed protein Omp10.
 XX
 KW Omp10; outer membrane protein 10; surface exposed protein;
 KW antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN M09858953-A2.
 XX
 PD 30-DEC-1998.
 XX
 PF 19-JUN-1998; 98MO-DK00266.
 XX
 PR 23-JUN-1997; 97DK-0000744.
 XX
 PA (BIRK/) BIRKELUND S.
 PA (CHRIT/) CHRISTIANSEN G.
 XX
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mybind P;
 DR WPI; 1999-105610/09.
 DR N-PSDB; X06822.
 XX
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 XX
 PS Claim 7; Page 60-62; 115pp; English.
 XX
 CC This polypeptide comprises the novel 98.4 kDa surface exposed
 CC protein Omp10 of the human respiratory pathogen Chlamydia
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see
 CC X068422) isolated from a C. pneumoniae expression library. The
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
 CC (see W88417-28), and nucleic acid sequences encoding them (see
 CC X06816-27). A new species specific test is claimed that is used
 CC to identify mammals (including humans) infected with Chlamydia

CC pneumoniae. The test compiles detecting antibodies specific for
CC OmpA-comp.5 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of *C. pneumoniae* infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines, for effecting *in vivo* expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with *C. pneumoniae*.

SQ Sequence 928 AA;

Query Match	41.68;	Score 1986;	DB 20;	Length 928;
-------------	--------	-------------	--------	-------------

Best Local Similarity -45.3%; Pred. NO. 1.7e-123;
Matches 430; Conservative 149; Mismatches 327; Indels 44; Gaps 19;

OY	1	MKSSEPKVFESTFAIFPLSM-----LATEVLDDSSAFSGDNKKNGNFSVRESQEDA-GTT	53
Dd	1	mksslhwlvjsssalalprlslnfsfaavevnlprntnsfsg--pdytvpqatnagdtl	58
OY	54	YLFGKWVLEMI PGNGALITKSCFNNTGDTLFFNGNSLSLFQTVDAGTAVAGAIVNSSVV	113
Dd	59	ynltgdvsitn-agspalatscfcetugnlsifghyqflgnidag--ancftntaa	115
OY	114	DKSTTFIOFSSLFTIASFSGSITTGGKAIVSCSTOSLSLTKNVSLFKSNFSTDNGAITA	173
Dd	116	nklisfgsfysljstlqt--tnattgtgaik-sigacsiqnyscyfqngtfndngalg	172
OY	174	KTLSTGTMTALFSENMSRKKGAIORSDMLITTGNOGEVSFDNTSPOGAIFEAFS	233
Dd	173	sstsis-Inpmlttaknkatkqgalyistvgitltnltncsasfentaaamgaalyeas	221
OY	234	VTISNNAKVSFIDNKVYTGA SSTTGDMSGAICAYKTSTDTKV-TLNGOMLFSNNTST	292
Dd	232	sfi snkai sctsi nsv tesa -----tgayicystsaprvltldngelnffigtai	285
OY	293	TAGGIAYKKKLASGGU TLFRSRVSGCNP RKKGAAIAIEDSGELS SASOGDIVLGNT	352
Dd	286	tsggallycdnlvissggpltkmnsaidtaaplggaataadsgelsistalgddtegn	345
OY	353	V---TSTPGINRSIDIG-TSARKMIALRASAGRAIFYDPITTGSSSTTVDLKVEVP	408
Dd	346	vkgasssqtltnsinignmnaivqlrasgntlyfybrpitnbhaalsdaalnpgd	405
OY	409	ADSA LQYTGNIITFGEKLSSETPAADSNNLSKLLQPYTSLSGGTSLKHNVTLOTQAFTQO	468
Dd	406	Iagnpaygtlivfegeklseaeaaadnktstloqrltaggqsltskyvtllvakstfsg	465
OY	469	ADSRLEMVDGTTLPADTSTINNIVINISSIDGAKKRIEFKATSKULTLSTGYTLLDP	528
Dd	466	pgrtllnadgctltadgitltnnlvinovdslekkckatlkatgasqcvltsgastlvdp	525
OY	529	GTFEYHSLRNPOSVIDLELKAS--GTVTSTA VTPDEIMGEKFHYXGYQTGWBPWOTGA	586
Dd	526	gvnyvedswmpgyfscelttladganhlitldaapbleknphlwpygmwa-lswgedt	584
OY	587	ST---TPTFMNTKICYIPNERISGLVPSNISMANFIDISSLHTLMETANGLGDPAFWC	643
Dd	585	atskaeatllwtcktygypnperrjclvaantlwsfvdvzvisiqglvalkvrgsqetrcwge	644
OY	644	AGLSNFPHKDSTRKRGRFRLSGGVYVGNLHTCCSDILIASACOLRGPRDRDFVAKNQG	703
Dd	645	eglsenlfhkdstkinkgrfnhsagyavgatllasdnlltaatecqglfgkrdrhfinkra	704
OY	704	TYVCGTLYQHNEITYISLPCKLRPCSLSYP---TELPUVFSGNLSTYTTDNDLKRKYTT	760
Dd	705	sayeaashlqlhalts-----psllitylpssesepvlidaqisyivsytkmfkytc	758
OY	761	YPTVKSGMGNDSPFLTEFGGRAP-ICIDESALFEQYMFMLOCVYAHOEFCROGNE-AR	818
Dd	759	apbgesswynogdcalleastsphtalsheglfhayfpfrlveasylybdgsfkernlttyr	818

[illegible]

RESULT

ID Y35052 standard; Protein; 930 AA.

AC Y35052;

DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae surface exposed polypeptide.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN W09927105-A2.

PD 03-JUN-1999.

PE 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

XX

XX
2
3
4
5
6
7
8

XX
DD
FBI : 1000-357643/20

XX
DE

XX
F200 CAG-013
7.6e
TGT388
F2714

XX
v3AE0A-v3E07C

frames in the complete genome (see X91990) of *Chlamydia pneumoniae*.

CC bronchitis and is thought to be a contributing factor in heart

CC frames of the C pneumoniae genome (500 X324584-X325879) can be used for nodosum or pharyngitis. The polypeptides encoded by the open reading

immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions

CC especially where the vector directs the expression of a neutralising epitope of *C. pneumoniae*.

AA
AA
Sequence 930 AA.

... SQ Sequence 930 AA:

Best Local Similarity

[illegible]

	—	7
	—	8
	—	9
	—	10
	—	11
	—	12
	—	13
	—	14
	—	15
	—	16
	—	17
	—	18
	—	19
	—	20
	—	21
	—	22
	—	23
	—	24
	—	25
	—	26
	—	27
	—	28
	—	29
	—	30
	—	31
	—	32
	—	33
	—	34
	—	35
	—	36
	—	37
	—	38
	—	39
	—	40
	—	41
	—	42
	—	43
	—	44
	—	45
	—	46
	—	47
	—	48
	—	49
	—	50
	—	51
	—	52
	—	53
	—	54
	—	55
	—	56
	—	57
	—	58
	—	59
	—	60
	—	61
	—	62
	—	63
	—	64
	—	65
	—	66
	—	67
	—	68
	—	69
	—	70
	—	71
	—	72
	—	73
	—	74
	—	75
	—	76
	—	77
	—	78
	—	79
	—	80
	—	81
	—	82
	—	83
	—	84
	—	85
	—	86
	—	87
	—	88
	—	89
	—	90
	—	91
	—	92
	—	93
	—	94
	—	95
	—	96
	—	97
	—	98
	—	99
	—	100

ト
ミ
ト
ト
ト
ト
ト
ト

	—	C	S
	—	T	S
	..	T	S
	—	T	S
	—	C	S
	—	C	S

[illegible]

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

```

Db      119 lftfmslfsiaapgtlvaaagkcls--sagalnldngllfsgnvsneannngaika 177
QY      174 KTLSTGTTMSALFSENNSSKKGAIQTSDALITINGOGEVSFSDNTSSDGAIFTEAS 233
      178 klslsgntssltfensakklygalaissaaslsqntqglvfmmhkgstggaglygfaas 237
QY      234 VTISNNKAVSFIDNKVTKGASSSTTGDMSGCAICAKTSTDTRVTLTGNMLFNNSTST 293
      238 sstltnssltfsgntatdaag-----xggaaycektgetpeltltsgnksltfaensvt 291
QY      294 AGCAIYVKKLELASGGLTFESRNSVNGTAPKGAIAIDSEGLSLASGDIVFGNTV 353
      292 gqgaicahglidisaagpflfsmrcgntaagkygalaidsaglsisaagdlitlgnl 351
QY      354 TSTT-PGTRSSIDTGTSAKMTALRSAGRAIFYDPI---TTGSSSTVTVDLKVNTERPA 409
      352 tsasptstrnalygssakltnlraagqgslyfydpilasnttgas----dvltnqps 407
QY      410 DSALOYTGNIITGKELSTEADSKNLTSLQPVLTSGTSLKHGVTLOTQAFTOQA 469
      408 nspldysgtlvisgeklasdeakaadnftslkqplalagelakgnvelvngftqte 467
QY      470 DSRLMDVGTTELEPADTSTIN--NLVINISIDGAKKAKETKATSKNLTSLGTTLLDP 527
      468 ggtllmqptkik-adteaistklivdlsalegnksvleagankltltsplvfqds 526
QY      528 TGTTFENSLRNPQSDIIELKAS-----GTVSTAIVPDPIMGEKHFHYOGTWG 578
      527 sgnfyeshlnaftcpqlvavftaataasdiyalaltspvqpep-----hygyqghwe 580
QY      579 PIWVGASTATFPMWTKGYIPNPERISLVPNSLWMAFIISLHYLMETANEOLOQD 638
      581 atwadstaksgtmtcvltqgynpnerisvvpdsfwasfdtrltqqlmtsagansiyq 640
Db      639 RAFWCAGLSNFHKDSTKRRCGRHLGGVYIGNLHTCSDKILSAFQOLFGRDRTV 698
      641 rglwasgtanftfhkdsngaftrhksygyivgssaedfsemlfsvafqglgkdxldlfi 700
QY      699 AKNCGVYGGTLYYOHNEHYISLPCLRPC--SLSYVPEIIVLFESGNSTYHTDLDLT 756
      701 ventshnylasylyqhratlggip---mpsfsgitdmldipllnaqlsystkdmmt 757
QY      757 KYTTYTVKSGWGNDSFALEFGGRAPICL-DESALEQOYMPMKLOFVYAHOGFEEQGT 815
      758 rlytspeagswtmsngalelsgslalykpeapffqgyfplfkfgavysrqnfkesga 817
Db      816 EKREGESSSLVNLALPIGRFDKESQCDATYNLTIGYVDLVRSNPDCTTLRLISGDSW 875
      818 eeattdagdlivncslpvqglrllekisedeknnfelslaylgdvyrknprrstlmvsgasw 877
QY      876 KTFGTNLARQALVLRAGNHFCFNSNFARSOSFELRGSSRNVNDLGAKYOF 928
      878 tscknlaraqalaasgshltlshpvelsgeaayelrsgahlynvdcglrlysf 930
Db
RESULT 10
W88424
W88424 standard; Protein: 930 AA.
XX
AC      W88424:
XX
DT      26-APR-1999 (first entry)
XX
DE      Chlamydia pneumoniae surface exposed protein Omp11.
XX
KW      Omp11; outer membrane protein 11; surface exposed protein;
XX      antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.
XX      Chlamydia pneumoniae.
XX      OS
XX      MO9858953-A2.
XX      PN
XX      30-DEC-1998.
PD

```

```

XX      19-JUN-1998; 98WO-DK00266.
PF
XX      23-JUN-1997; 97DK-0000744.
PR
XX
PA      (BIRK/) BIRKELUND S.
PA      (CHRIV/) CHRISTIANSEN G.
PI      Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI      Mygind P;
XX      WPI: 1999-105610/09.
DR      N-PSDB: X06823.
DR
XX
PR      Species-specific test for identifying mammals infected with
PR      Chlamydia pneumoniae - comprises detecting antibodies specific for
PR      outer membrane proteins of C. pneumoniae or nucleic acids encoding
PR      these proteins
XX
XX      Claim 7; Page 63-65; 115pp; English.
XX
CC      This polypeptide comprises the novel 97.6 kDa surface exposed
CC      protein Omp11 of the human respiratory pathogen Chlamydia
CC      pneumoniae. Its amino acid sequence was deduced from DNA (see
CC      X06823) isolated from a C. pneumoniae expression library. The
CC      invention provides 12 novel surface exposed proteins, Omp4-Omp15
CC      (see W8417-26), and nucleic acid sequences encoding them (see
CC      X06816-27). A new species specific test is claimed that is used
CC      to identify mammals (including humans) infected with Chlamydia
CC      pneumoniae. The test comprises detecting antibodies specific for
CC      Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC      membrane proteins, especially by PCR. The proteins are also used
CC      in the diagnosis of C. pneumoniae infection in mammals. The
CC      nucleic acids and proteins can also be used in the immunization of
CC      mammals, the nucleic acids being particularly useful as DNA
CC      vaccines for effecting in vivo expression of antigens. The
CC      vaccines may also prevent atherosclerosis and bronchial asthma,
CC      which are possibly associated with C. pneumoniae.
XX
XX      Sequence 930 AA:
SO
Query Match 40.4%; Score 1927; DB 20; Length 930;
Best Local Similarity 43.4%; Pred. No. 1,36-119;
Matches 414; Conservative 165; Mismatches 326; Indels 48; Gaps 17;
QY      1 MKSSFPKVPSTFAFPISM-IAT---ETVLDSSAKSPFGKNKNGNSVSESOEDACTITLF 56
      1 mkpdlkllssltvtplllslatygadaaslpdsldagsgstlpxstadaangtnyvl 60
Db
QY      57 KGNVLENIIPGTAITKSCFNNTSGDLPFGNGNSLFFQVVDAGTVGAAVNSVVS 116
      61 sgnvyi-ndagktalycctfletqdlitfgkyrsfntvdagsnagaaa-stdacka 118
Db
QY      117 TTFIGFSSLSFASPSSITTGKAVSCSTGSLSTKKNVSLFSKNFSTD---NGCAITTA 173
      119 lftfmslfsiaapgtlvaaagkcls--sagalnldngllfsgnvsneannngaalt 177
Db
QY      174 KTLSTGTTMSALFSENNSSKKGAIQTSDALITINGOGEVSFSDNTSSDGAIFTEAS 233
      178 klslsgntssltfensakklygalaissaaslsqntqglvfmmhkgstggaglygfaas 237
Db
QY      234 VTISNNKAVSFIDNKVTKGASSSTTGDMSGCAICAKTSTDTRVTLTGNMLFNNSTST 293
      238 sstltnssltfsgntatdaag-----xggaaycektgetpeltltsgnksltfaensvt 291
Db
QY      294 AGCAIYVKKLELASGGLTFESRNSVNGTAPKGAIAIDSEGLSLASGDIVFGNTV 353
      292 gqgaicahglidisaagpflfsmrcgntaagkygalaidsaglsisaagdlitlgnl 351
QY      354 TSTT-PGTRSSIDTGTSAKMTALRSAGRAIFYDPI---TTGSSSTVTVDLKVNTERPA 409
      352 tsasptstrnalygssakltnlraagqgslyfydpilasnttgas----dvltnqps 407
Db

```


Db 755 ytgpkqgsswyndgcallealsiphtalsheglfhayfifkweasyihbdsfkerntt 814
 QY 817 -AREFGSSRLVNLALPIGIRFDKESDCODATYNTLGYTVDLVNSNPDCITTLRISGDSW 875
 Db 815 lvsifdsqdlivnvspigltfiferfsernerasyeatviyvadvyrknpdcttalllnntsw 874
 QY 876 KTFGTNLARQALVLRAGNHFCFNSFEAFSQSPFELRGSNNYVNDGAKQOF 928
 Db 875 ktgtnlstrqgigraglifaifsplevtnlsmeltrgsrsynadlgykltgf 927

RESULT 12

Y35060
 ID Y35060 standard: Protein: 949 AA.

XX Y35060:

DT 13-SEP-1999 (first entry)

XX Chlamydia pneumoniae cellular envelope protein.

XX Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KM vaccine; neutralising epitope.

XX Chlamydia pneumoniae.

PN W09927105-A2.

XX 03-JUN-1999.

PE 20-NOV-1998: 98WO-IB01890.

XX 04-NOV-1998: 98US-0107078.

PR 21-NOV-1997: 97FR-0014673.

XX (GENSET) GENSET.

PI Griffiths R;

XX WPI: 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

XX Page 947-949; Disclosure: 1912pp; English.

XX Y34584-Y35879 represent the proteins encoded by all the open reading

CC frames in the complete genome (see X91990) of Chlamydia pneumoniae.

CC C. pneumoniae causes respiratory disease such as pneumonia and

CC bronchitis and is thought to be a contributing factor in heart

CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema

CC nodosum or pharyngitis. The polypeptides encoded by the open reading

CC frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in

CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae

CC nucleotide sequences can also be used as immunogenic compositions,

CC especially where the vector directs the expression of a neutralising

CC epitope of C. pneumoniae.

XX Sequence 949 AA:

Query Match 38.9%; Score 1856; DB 20; Length 949;

Best Local Similarity 42.7%; Pred. No. 7,1e-115;

Matches 405; Conservative 171; Mismatches 331; Indels 42; Gaps 20;

QY 1 MKSSPFKVFVSTFAIFP--LSMATEVVLDSASFQDN-KKGNFVSRESQDCACTTLFLK 57

Db 22 mktisipwlvsvsvalafschlqslaneellspddsfngnidsgftfpkts---atyslt 77

QY 58 GNVLTLENIPGTGTAITKSCFNNTKGLFTFGNSLLFGQTVADAGTVGAAVNSVVDKST 117

Db 78 gdvffye-pgkqclpsdscfktctdnltlfignhsiltgfidagcthgagaa-stantknl 135

QY 118 TTFIFSSLSLTASPCSSITTTGKCAVSCSTGSLSTTKVSLFLFSKNFTDNGAITAKTLS 177
 Db 136 tfsgfslstidspsltvtvgqtlis--sagvnlennlrklvngnfstadgaikgasfl 194
 QY 178 LCTGTMALSEMNSSKGAIOIOTSDALTTTGNQGEVFSFDNMSDSCGAIFTEASVTIS 237
 Db 195 ltgysgallismssckkgalatltagariantqxrflsnlaasgsaiddeslils 254
 QY 238 NNAKVSFIDNRKVTGASSTTGDMSSGGAICAYKTSTDTRKVLITGNOMLFSSNNTSTAGGA 297
 Db 255 nktflyf-----egnaekt-----ggaicntkasgspeellismnktliffsnvaetsgga 305
 QY 298 IYVAKLELASGGLTFERNVNGTAPKGAIAIEDGELSADSQDYLFLGNTVST- 356
 Db 306 ihakklaissggtfeelfrmvssat--pkgaaisdaagselisaetnltfvrnltltg 364
 QY 357 -TPGTRSSSIDLGTSAKMTALRSAGRAIFYPIPTGSSSTVTVDVKVEPTADSALOY 415
 Db 365 stdtkrtnainisngkfiteiraaknhtliffydpit--segtsdvtkingasaalmpy 422
 QY 416 TGNITFTGEKLETEAADSNNLTSKLLQVYTLGGTSLKHGVTLOATQOADSRLKM 475
 Db 423 qgtllfsgeltadelkvadnlkssftqpvslsgklllkgyvlestsfsgaagsligm 482
 QY 476 DVGTTLE-PADTSTINNLVINISSIDGAKKAKIETKATSKNLISGHTITLDPRTFREN 534
 Db 483 dsqctlstagsititcniinvdsglqkpvslitakyaasnkvlsyglnlidlegniyes 542
 QY 535 HSLRNPQSYDILELKASGTVTS---TAVTPDPIMEGFHYGQGTWGPVMTGASTT- 589
 Db 543 hmfshdqllfalkitvdadvdtnvdisslipyaeodpenseyfgqgvn--vnwtctlatnt 601
 QY 590 --ATFNWTKTGYPNPERISLVPNSLWNAFIDISLHYMETRANBSLQDRAFWCAGLS 647
 Db 602 kealatwktgtyvpsperkalvcnltwgyftldrsldqiveigatcmehkqgfwvssmt 661
 QY 648 NPFHKDSTKTRGRFHLGSGYVIGNLTCTSDKILSAFQLFGDRDYPVANKOGTVYG 707
 Db 662 nfhktgdenrkgftrhsagvyigsahtpkdclftafchlfarddciaannsttyg 721
 QY 708 GTLYYOHNET----YISL-PCKLRPCSLSYVPEIPVLFSGNLSYTHDNDIKTKYTYT 761
 Db 722 gtlffkshltqpnlylrigrakfsesalekfprelpladvgvsfshsdnrmethysl 781
 QY 762 PTYKSGWGNSPALFEGGRAPICL-DESALEBOYMPMKLOFYAHQEGKEGTEAREF 820
 Db 782 peesegswsncisagglldlplfvlsnphlftclpkmkvemyvasqnsfessdgrgtf 841
 QY 821 GSSRLVNLALPIGIRFDKESDCODA--TYNLTGYTVDLVNSNPDCITTLRISGDSWKTFG 879
 Db 842 sigrlnlslspgakf--vgqldgdslydlsigfsvdyvynnpqstactllwmspdswwkig 900
 QY 880 TNARQALVLRAGNHFCFNSFEAFSQSPFELRGSNNYVNDGAKQOF 928
 Db 901 gnlsrqaflilrgsnmyynsnancelfghyamelrsgsrnyvadvgtklrff 949

RESULT 13

W88417
 ID W88417 standard: Protein: 928 AA.

XX W88417:

XX 26-APR-1999 (first entry)

XX Chlamydia pneumoniae surface exposed protein Omp4.

DE Omp4; outer membrane protein 4; surface exposed protein; antigen;

KW infection; diagnosis; vaccine; atherosclerosis; asthma.

XX Chlamydia pneumoniae.

PN W09858953-A2.

xx 30-DEC-1998.
pd
xx
xx 19-JUN-1998; 98WO-DK00266.
xx
xx
xx 23-JUN-1997; 97DK-0000744.
xx
xx
xx (BIRK/) BIRKELUND S.
xx (CHR/) CHRISTIANSEN G.
xx
xx
xx Birkelund S, Christiansen G, Knudsen K, Madsen A;
xx
xx Mygind P;
xx
xx WPI: 1999-105610/09.
xx
xx N-PSDB: X06816.
xx
xx
xx Species-specific test for identifying mammals infected with
xx Chlamydia pneumoniae - comprises detecting antibodies specific for
xx outer membrane proteins of C. pneumoniae or nucleic acids encoding
xx these proteins
xx
xx Claim 7; Page 40-42; 115pp; English.

CC This polypeptide comprises the novel 98.9 kDa surface exposed
CC protein Omp4 of the human respiratory pathogen *Chlamydia*
CC pneumoniae. Its amino acid sequence was deduced from DNA (see
CC X06816) isolated from a *C. pneumoniae* expression library. The
CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
CC (see W88417-28), and nucleic acid sequences encoding them (see
CC X06816-27). A new species specific test is claimed that is used
CC to identify mammals (including humans) infected with *Chlamydia*
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of *C. pneumoniae* infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting *in vivo* expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with *C. pneumoniae*.

Sequence 928 AA;

Query Match	38.9%	Score 1855;	DB 20;	Length 928;
Best Local Similarity	42.7%;	Pred. No. 8e-115;		
Matches 405; Conservative	111;	Mismatches 331;	Indels 42;	Gaps 20;

[illegible]

Db	344	scdtpkrrainingsngkfctfelraaknhtlffdydlt--segtssdwlkinngsaglnpy	401
Qy	416	TGNIFTEGKLSFTEAASDKNLTSEKLDLPVLSGGTISLKHGVTLOQATFOCADSLEM	475
Db	402	gqtlifsegltladelkvadnlksftlpavlsqgklllqkvlelssfsgeasllgm	461
Qy	476	DVGTTLF-PADSTINNVINISSIDGAKKAKIEFKATSKULTSTGITTLLDPGTFFEN	534
Db	462	dsqctltstagsitftlnlginvdslyglkqpslckagasnkvlysglnlidiqnlays	521
Qy	535	HLRNPQSDIIEELKASCTVTS---TAVPDPIMGEKFKHYOGTGWPIWGTGASTT-	569
Db	522	hmfsndqflslklktlvdaadvtnvdissllypaeapdnseyfgfgvwn-vnwtldatnt	580
Qy	590	--AFENMWTGTVIMPPEKISGLVNSLNMNAPIDISLHYMETANEGLQDRAFWCGLS	647
Db	581	keatatktgtvpspekalsavcnltlwgvtfdlrslyqqlveigaqemhkgqltwssmt	640
Qy	648	NEFHKDSKTRKGRFHLISGCVYVIGNLTHTCSDKLLSAFCOLEGRDDEYFAKNOGTVYG	707
Db	641	nflhtkgdenkrgfhtsgyvyigsahtpbdllftafchlfardkdcflahnnstlyg	700
Qy	708	GLTYQNHET-----YLSL-FCKLRPCSLSVPEIPIEVLESNGNLSYHTNDLKTXTTY	761
Db	701	glflfkthhtclqpqnlylllgrakfsesalekfprelpaldavqvsfshsdnrmethylsl	760
Qy	762	PIVKGSGMGNDSFALFEFGGRAPICL-DESALEQOYMPMKLOFYAHOEGFEQGTAREF	820
Db	761	pesegswsneclaggigldlpfvtlsnphplfkftlpgmkvemyvsgnsffessdgrgf	820
Qy	821	GSSRLVNLALPIGRFPEKESQCDQA-TYNTLTGTYVDLVRSNPDCTTLRISGDSWKTFG	879
Db	821	sigrlrlnlslpygakf-qvgqldgdslytdsigffvsdvyrnnpqstatlvmsspdkrlrg	879
Qy	880	TNLAAQALVLRAGNHEFCNSFEKFSQFSFELRGSSSNVAVDCAKQOF	928
Db	880	gnlsrqalilrgsnmyynsnceflfghyamelilgssstrnyndvgkrlrf	928
RESULT 14			
ID	Y69368		
AC	Y69368	standard; Protein; 945 AA.	
DT	19-JUN-2000	(first entry)	
DE	Amino acid sequence of the CPN100394 polypeptide.		
XX	CPN100394; Chlamydia infection; immune response; vaccine.		
OS	Chlamydia pneumoniae.		
XX	Key	Location/Qualifiers	
FT	Peptide	1..43	
FT		/note= "signal peptide"	
PN	WO200011183-A2.		
PD	02-MAR-2000.		
PF	18-AUG-1999;	99WO-IB01449.	
XX	20-AUG-1998;	98US-0097187.	
PR	20-AUG-1998;	98US-0097188.	
PR	20-AUG-1998;	98US-0097189.	
PR	20-AUG-1998;	98US-0097190.	
PR	20-AUG-1998;	98US-0097195.	
PR	20-AUG-1998;	98US-0097196.	
PR	20-AUG-1998;	98US-0097197.	
PR	27-AUG-1998;	98US-0097191.	
PR	17-AUG-1999;	99US-0376770.	


```

XX (CONN-) CONNAUGHT LAB LTD.
PA
PA Mordin AD, Oomen RP;
PI
XX MPI: 2000-224703/19.
DR N-PSDB; 261508.
XX
PT Novel antigens and corresponding DNA molecules that can be used to
PT prevent, treat and diagnose disease caused by Chlamydia infection in
PT mammals, especially humans.
XX
PS Claim 18: Fig 13A-F; 201pp; English.
XX
CC Y69362-69 represent Chlamydia pneumoniae polypeptides. The polypeptides
CC are present in the bacterial membrane structure, in the external
CC vicinity of the membrane structure, in the inclusion membrane
CC structure, in the external vicinity of the inclusion membrane
CC and in the cytoplasm of the infected cell. The polypeptides may be
CC used to prevent, treat and detect the presence of Chlamydia infection
CC and/or the presence of Chlamydia in a sample. The polypeptides may
CC also be used to induce an immune response in a mammal. The vaccine
CC vector comprising the polynucleotides is used to induce an immune
CC response in a mammal. Antibodies directed against the polypeptides
CC may also be used therapeutically to treat and/or prevent a Chlamydia
CC infection.
XX
SQ Sequence 945 AA:

```

```

Query Match 38.4%; Score 1832; DB 21; Length 945;
Best Local Similarity 43.6%; Pred. No. 2,7e-113;
Matches 410; Conservative 151; Mismatches 328; Indels 52; Gaps 23;

```

```

OY 588 TTAATPWTKTGYPINPERIGSLVPSNLWNAFIDISSLHLMETANEGLOGDRAFCAGLS 647
DB 608 sganlewrtgylpnperqslvpsnlwsgsfvdqraqlqelmvnssqllcgergvwagla 667
OY 648 NFFHKDSTKRRGRFRLHSGGVYVGNLHNCSDKILSAFCOLGFRDRDFAVNAOGTVG 707
DB 668 nflhrdki-nehgyrinsgyvlgvgtlhaftsdaatnaicqltrdkdyvsnhgtys 726
OY 708 GTLYYOHNETYISLPCKLRCSLSYVPEFLPVLFSGNLSYTHDNDLKTYYTPYVKS 767
DB 727 gvfledtlefrs-pqgfydssseacnqvrlmdqslsyshndmkktytlypeags 785
OY 768 WGNDSFALFERGGRAPICLDSALFEOYMPMKQFYAHOBEGKEGTAREGSSRLVN 827
DB 786 wandvlglegfatlyypnsflfdyyspflrlqcyahbedikegyevrhtsgdlin 845
OY 828 LALPIGRPKESDCODATYNTLGYVDLVRSNPCTTLRLISGDSMKFTGNLARQAL 887
DB 846 lavpigvkferfdckrsgyeltlayvdrirkdpstall-asgatwethgmnlstrgl 904
OY 888 VLRAGNHFCSNFEAFSOPFELRGSSRNRYNDLCAKIOF 928
DB 905 qlrlghcllnpglevfshgatelrgsrnylnlgkyrf 945

```

RESULT 15

W88428 ID W88428 standard; Protein; 945 AA.

AC W88428;

DT 26-APR-1999 (first entry)

DE Chlamydia pneumoniae surface exposed protein Omp15.

KW Omp15; outer membrane protein 15; surface exposed protein;

XX antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.

OS Chlamydia pneumoniae.

PN W09858953-A2.

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK00266.

PR 23-JUN-1997; 97DK-0000744.

PA (BIRK/) BIRKELUND S.

PI (CHR/) CHRISTIANSEN G.

PI Birkelund S, Christiansen G, Knudsen K, Madsen A;

PI Mygind P;

DR MPI: 1999-105610/09.

DR N-PSDB; X06827.

XX Species-specific test for identifying mammals infected with

PT Chlamydia pneumoniae - comprises detecting antibodies specific for

PT outer membrane proteins of C. pneumoniae or nucleic acids encoding

PT these proteins

XX Claim 7: Page 71-73; 115pp; English.

CC This polypeptide comprises the novel surface exposed protein Omp15

CC of the human respiratory pathogen Chlamydia pneumoniae. Its amino

CC acid sequence was deduced from DNA (see X06827) isolated from a C.

CC pneumoniae expression library. The invention provides 12 novel

CC surface exposed proteins, Omp4-Omp15 (see W88417-28), and nucleic

CC acid sequences encoding them (see X06816-27). A new species

CC specific test is claimed that is used to identify mammals

CC (including humans) infected with Chlamydia pneumoniae. The test

Fri Nov 24 13:49:56 2000

us-09-428-122-2.rag

Page 15

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2000, 08:19:06 ; Search time 35.79 Seconds
(without alignments)
434.601 Million cell updates/sec

Title: US-09-428-122-2

Perfect score: 4774

Sequence: 1 MKSSFPKVFSTFAIFPLSM.....FELRESSRNYNDLGAKYQF 928

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/PCRTUS_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271.5	5.7	1338	2	US-08-728-470-9
2	271.5	5.7	1599	2	US-08-617-697-9
3	267	5.6	1529	2	US-08-728-470-10
4	266	5.6	1600	2	US-08-617-697-10
5	244.5	5.1	1026	2	US-08-614-377A-7
6	240.5	5.0	1536	1	US-08-038-682-2
7	240.5	5.0	1536	1	US-08-302-832-2
8	240.5	5.0	1536	2	US-08-302-832-2
9	240.5	5.0	1536	2	US-08-469-880-2
10	240.5	5.0	1536	2	US-08-728-470-2
11	240.5	5.0	1536	2	US-08-617-697-2
12	238.5	5.0	1026	1	US-08-194-290-7
13	234	4.9	1612	1	US-08-169-927-2
14	223.5	4.7	1477	1	US-08-038-682-4
15	223.5	4.7	1477	1	US-08-302-832-4
16	223.5	4.7	1477	2	US-08-530-198-4
17	223.5	4.7	1477	2	US-08-469-880-4
18	223.5	4.7	1477	2	US-08-728-470-4
19	223.5	4.7	1477	2	US-08-617-697-4
20	202.5	4.2	674	1	US-08-317-522A-3
21	202.5	4.2	674	1	US-08-439-818A-3
22	202.5	4.2	674	2	US-08-751-965-3
23	202.5	4.2	674	2	US-08-738-975-3
24	202.5	4.2	674	2	US-08-728-626-3
25	202.5	4.2	674	3	US-08-808-599A-3
26	202	4.2	1912	1	US-08-409-995-4
27	202	4.2	1912	3	US-08-685-467-4
28	201.5	4.2	1160	3	US-08-808-599A-24

29	200.5	4.2	749	1	US-08-317-522A-2	Sequence 2, Appl
30	200.5	4.2	749	1	US-08-439-818A-2	Sequence 2, Appl
31	200.5	4.2	749	2	US-08-751-965-2	Sequence 2, Appl
32	200.5	4.2	749	2	US-08-738-975-2	Sequence 2, Appl
33	200.5	4.2	749	2	US-08-728-626-2	Sequence 2, Appl
34	200.5	4.2	749	3	US-08-808-599A-2	Sequence 2, Appl
35	181.5	3.8	1248	2	US-08-348-353-17	Sequence 17, Appl
36	181.5	3.8	1248	2	US-08-465-965-17	Sequence 17, Appl
37	181.5	3.8	1248	3	US-08-465-966-17	Sequence 17, Appl
38	169.5	3.6	983	2	US-08-164-292B-26	Sequence 26, Appl
39	169.5	3.6	983	3	US-08-845-623-26	Sequence 26, Appl
40	169.5	3.6	983	3	US-08-845-623-26	Sequence 26, Appl
41	164.5	3.4	1222	2	US-08-682-517-15	Sequence 15, Appl
42	164.5	3.4	1222	2	US-08-682-517-9	Sequence 9, Appl
43	162	3.4	1394	4	PCT-US95-1061A-2	Sequence 4, Appl
44	160.5	3.4	666	2	US-08-737-716-14	Sequence 14, Appl
45	159	3.3	907	3	US-08-783-774-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-728-470-9
; Sequence 9, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berktessier, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0810
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-728-470-9

[illegible]

FILING DATE: 01-APR-1996
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/302,832
 FILING DATE: 05-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US PCT/US93/02166
 FILING DATE: 16-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Berkstresser, Jerry M
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-557
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1599 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

Query Match	Similarity	5.7%	Score 271.5	DB 2:	Length 1599:
Best Local	Similarity 22.5%		Pred. No 2.1e-12:		
Matches 146:	Conservative 81:		Mismatches 216:	Indels 205:	Gaps 29:
QY	23	TETVDSASPEGNKNGFVSRES-----	QEDAGTYY--LFGKVTLENI	65	
DB	912	TVNTEVAVGSEFPNNCASNISIRGAKFKEDINNTSSLNTITNTSDTYRILIIKGNIS--NK		969	
QY	66	PETGTAIRKSC-----	FNNTGDLTFENGNGSLLFQ--TVDGATVAGAAVNVSVDK	115	
DB	970	SDDLNIIDKKSDAEIQIGGNIISQKGNLIISDKXNIITQITIKKAGVEGRDSSAEANA	1029		
QY	116	STT-----	FIQFSSLSFASBPSSITTKGAVSCSTGSLTK-----	153	
DB	1030	NLIQTKEKLKLAGDLINISGFNAKEITAKKNGSLITIG----	NASGNADAKKATYFDKVKDS	1085	
QY	154	-----	NVSLFSKNEFSINDGCAITAKTLLSLGTMSA--LFSENNSSK-----	G 196	
DB	1086	KISTGDHNTL--	NEEVKTSNGSS--NAGNDNSTGLIKSKADVTNNNNVTSHTKINISAAA	1142	
QY	197	GAIQTSDALITIGNGGEVSEFSDNTSSDGAAFTEASVYISNNAKVSFIDNKVYTGASSST	256		
DB	1143	GNVTKEGTTIATNGTSVEVTAQNGIKGN--ITSQNVATFENLVTEENAVINATFST	1200		
QY	257	-----	TGDMSCGAICAVTSTDTKVYTLTGOMLLEFSNNT-----	STAGGA 297	
DB	1201	VNISTKTDGIKGG--	TESTSGNVNITTAGGN--TLKVSNTIGODVYATADAGALLTTAGST	1256	
QY	298	IYVK-----	KLEIASGGLTFERNVSNGTAFKGAIAI--EDSGEL----	337	
DB	1257	ISATIGNAIIITKTGDIINKVESSSGSVTLVATGALVANGINSGNVTIITADSGKLTSTV	1316		
QY	338	-----	SLASDGGI--VFLGNTVYSTTPGTRRSSIDLTGSAKKMALASAARAIY	385	
DB	1317	GSHINGTNSVITSSOSGDIEGTIGCNTVAVT--	ASTGDLITGNSKAKVAKNGAA-----	1368	
QY	386	FYDPITGTSSTVTVDLKVNENFPADALQYTGNIIFTEGKLESETAEADSKNLTSLKLOPV	445		
DB	1369	-----	TLAESGKLTITQIGSSI	1385	
QY	446	TLSGG--TLSLKH--	GVTLQIQ--AFTQADSRLEMDVCT--TLEPAD-----	T 486	
DB	1386	TSSNGCOTTLAKKSSIAGNINANAVTLNTGTLTTGGDSKINATSGTLTINAKDARLDGA	1445		
QY	487	STINNLVINISSIDGAKKAKIETKATSKULLISGTLTDPGTFEVN	534		
DB	1446	ASQDRIVAVNATNAGSGGNV--	TAKTSSSVNITGDLNTINGNLITSEN	1490	

Query Match 5.6%; Score 266; DB 2; Length 1600;
Best Local Similarity 22.6%; Pred. No. 5.7e-12;
Matches 147; Conservative 75; Mismatches 215; Indels 214; Gaps 29;

```
OY 31 ASFDGNKNGNFSVRESQEDAGTYYL-FKGNVTL-----ENIPGTGATTKSCF 77
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 908 STFKGEASDNLNLTGTFTNGTANIKGVKLGIDINNGKGINITNMGQKTIINGNI 967
OY 78 NTKGDLFT-----GN-----GNSLLFQ-----IYDAGTVAGAAVNSV 112
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 968 TNEKDDLNIKIKADAELQIGNISQKEGNLITSSDKVNITNQLIKAVVEGGRSDSEA 1027
OY 113 VKSTT-----FIFGSLSLFASPGSSITTKGAVSCSTGSLTLK----- 153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1028 ENANLTIQTKELKLAGDLNISGFNKAETAKNGSDLTIG-----MAGGNADAKKATFPDKV 1083
OY 154 -----NVSLFSKNFSDNGAATPAKTLTLGTTMA--LFESENTSKR----- 195
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1084 KDSKISTDGHVTL--NSEVKTNGSS--NAGDNSTGLTISAKDVTYNNNVTSKHTINIS 1140
OY 196 --GGAIQSDALITINGCEVFSFSDTSSDGAALFTEASVTISNNAKVSFDKNVTGAS 253
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1141 AAGAVTTKEGTTINATGTSVETLQNGTIKGN--ITSONVTVTATENLVTTENAVINAT 1198
OY 254 SST-----TGDMGSGAICAYKSTDTKVTLTGNOMLFSNNT-----STTA 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1199 SCTVASTKGTGDIKGG---IESTSGVNTITAGSN--TLKXSNITGQDVYVTLADAGLITTA 1254
OY 295 GGAIVK-----KLELASGGLTLFSRNSVNGGTPARKGAIAI--EDSGEL- 337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1255 GSTIATGNANITTKTGIDINGKESSGCVTLVATGATLAVGINSGMTVTTADSGKLT 1314
OY 338 -----SLSAASGDI--VELGNTVSTPGTNRSSIDICTAKMALLSAAGR 382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1315 STVGSTINGNTSVTSSGSDIEGTISGTVAVT--ASTGDLTIGNSKAEKAGAA-- 1369
OY 383 AIFYDPITTGSSITVTLVKNETPADALQYTGNIITFGKLSLSETEADSKNLTSKLL 442
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1370 -----TLTAEGRKLTITGTG 1383
OY 443 QPVTLISG--TILSKH-----GVTLQIQ-AFTQOADSRLMDVGT-TLEPAD--- 485
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1384 SSISSNQOTLTIAKDSSIAGININANVTLNTTGTTLTTGDSKINATSGTLINAKDAKL 1443
OY 486 --TSTINNLVINISSIDGKKAKIETKATSKVLTLSGTLTLLDPGTIYEN 534
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1444 DGAAAGDRTVAVNATVASSGAV--TAKTSSSVNITGDLNTINGLNTISEN 1491

RESULT 5
US-08-614-377A-7
: Sequence 7, Application US/08614377A
: Patent No. 5978864
: GENERAL INFORMATION:
: APPLICANT: Smith, John
: APPLICANT: Bingle, Wade H.
: APPLICANT: No. 5978864ellini, John F.
: TITLE OF INVENTION: EXPRESSION AND SECRETION OF
: TITLE OF INVENTION: HETEROLOGOUS
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson PC
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,377A
FILING DATE: 12-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/194,290
FILING DATE: 09-FEB-1994
CLASSIFICATION: 435
APPLICATION NUMBER: US 07/895,367
FILING DATE: 09-JUNE-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Tsao, Y. Rocky
REGISTRATION NUMBER: 34053
REFERENCE/DOCKET NUMBER: 08106/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1026 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-614-377A-7
```

Query Match 5.1%; Score 244.5; DB 2; Length 1026;
Best Local Similarity 24.2%; Pred. No. 1.4e-10;
Matches 179; Conservative 86; Mismatches 257; Indels 217; Gaps 33;

```
OY 12 TPAIFPLSMITEVLYDSSASFQGNKNGNFSVRESQEDAGTYYL-FKGNVTLNIP---G 67
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 TFVAGVAGAAATLVGDTLISGAGT---DVLNMGVAAAVTALPTG-VITISGIETNMVT 322
OY 68 TGTATKSCFNNKTGDLFTFTGNGSLLFQYVDAG-----TVGAAVNNSVDKSTFI 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 SGAATLTLTSSGVTG-LTALNTNTSGAAQVYTAGAGNLTTATAQAANNVAVD----- 375
OY 121 GFSSLSFTASPGSSITTKGAVSCSTGSLTLKRVSLFSKNFSTDNCGAITYATLSLTG 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 GRANVTASTGVTSGTTVGANSAAAGTSV-----SVANSTTTGAIATVG 423
OY 181 TTMALPSENISKKKGAIQISDALITINGNG--EVSFSDTSSDGAALF--TEASVTI 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 GT--AVTVAQTAGAANVTTLTQADVTYVGNSTTAVTVTQTAATAGATVAGRVNGAVTI 481
OY 237 SNNA-----KVSFIDNKVTGA-----SSSTGDMSG-----GAICAKYKSTDI 274
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 482 TDSAAASATTKGATATYVLGSGAATIDSSALITVINLSGTGTSIGIRGALITA----- 534
OY 275 KVLTLGNOMLFSNNTSTTACGAIYVKKLELASGTLTFSRNSVNGGTPAPKG-GAIAIED 333
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 535 --FTANTLTLNVNGLTFT--GALTDEAAADG--FTTINAGSTASTIASLVAAD 586
OY 334 SGELSLSDSDIV-----FLGNTVYTS-----TGTNRSSIDELGSA 371
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 587 ATTINISGDARVTTTSHTAALTGITVNSVATLGAELATGLVETGGAGRDSILLGAT 646
OY 372 KMTALRSAGRAIFYPIPTTGSSTV-----TDVLKVN----- 405
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 647 K--AIVWAG-----DDTYVSSATLGAAGSVNGGCTDVLVANVNGSSFSADPARGCF 698
OY 406 ET--PADALQYTGNI--FTGEKLSETEAA----- 432
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 699 ETLRVAGAAAGSHNANGFTALQIGATAGATFTTNVAVNGLVLAAPGTGTTVTLANAT 758
OY 433 ---DSKULTSKLQPVLLSGTSL-----KRGVTLQTA-----FT 466
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 759 GTSDVFNLTL--LSSSALAAGVYALAGVETVNIATDTNTTAHVDLTLLQATSAKSIYV 816
```


QY 467 QOADSRLMDVGTLEPADSTI-----NNLVINISSIDAKAKIETKATSK 514
DB 817 GNGGLNLTNGNTVSEFSDASAVGTAPAYTVFSANTVEVYIRGAGADSLTGSATA 876
QY 515 NLTL-----SGTITLDPGTGYENHSLRNPOSYDLLEKASGVSTAVFTPDPIGSEFHY 571
DB 877 NDTIIGAGADTLVYTGCT-----DTFTGIGADIEDFALNIGTSTAFVTTIDAAGDLDL 932
QY 572 GYGCTWGPVWGT-GASTT 589
DB 933 VGISTNGAIDAGAFGAAYT 951

RESULT 6
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-038-682-2

Query Match 5.0%; Score 240.5; DB 1; Length 1536;
Best Local Similarity 22.6%; Pred. No. 5.3e-10;
Matches 139; Conservative 106; Mismatches 220; Indels 149; Gaps 29;

QY 20 MATEVLDSSAFDCKNKGNSVRESQEDAGTTLFKGVLTLENIPGTATATKSCFN 79
DB 818 MICKGVAKKNITFEGBG-NITFEGRKAYTE-----IEGVNTINN--NAVVTILGSDFDN 868
QY 80 TKCDLT-----FTGNGNSLLFPYDACTVAG-AAVNSSVVDKSTFTIGFSSLSF 127
DB 869 HOKPLTIKKDVIINSGLTAGGNIV-----NIAGNLVESNANFKAITNFTFNWGL 920
QY 128 IASPG-SSITTGKAVSC-----STGSLSLTKNS-----LLFSKNSFDNGCATATKLSL 178
DB 921 FDKKNSNISIAKGARFCDIDNSKNLSTTSSSTYRTIISGNITKNG-----DLNT 974

QY 179 T--GTTMSALFSENNTSSKKGAIQTSDALTT-----GNOGEVSEFSDNTSSGCAIIF 229
DB 975 TNEGSDTEMQIGGDVSOKEGNLTISSDKINITKQITIKAGVDGENDSDATINNANLTIKT 1034
QY 230 TEASVT-----IS--NNAKVSFDNNKVTGASSSTTGMGSAICAYTSDTKYTLGNOM 283
DB 1035 KELKLTQDLNLSGFNKAETAKDGSDDLGTGNTNSADGTAKRYVNOVYDKSLISADGHV 1094
QY 284 LIFS-----NNTSTTGAIAIYVKKLELASGGLTIFSRL-SVNGCTAPKGAIAIEDS 334
DB 1095 TLHSKVETSGSNNTTEDSSDN-----NGLIITDAKNYVNNN-----ITSH 1135
QY 335 GELSLSADSGDIYF-IGNVTYSTTPG---TNRSSIDLG---TSAKMTALRSAGRAIYF 386
DB 1136 KAVISATSGEITTKGTITINATGNVEITAGTGSILGESSGSGVTLTATGALAV-- 1193
QY 387 YDPIITGSSSTVTVDLVKVNTPADSALOYTGNITFTGEMKSEFSEADSKNLTSLQPT 446
DB 1194 --SNISGNVTVTANGALTTLLAGSTIKGTESV-----TSSQSGDIG-----T 1236
QY 447 LSGGTLSLRHGVTLQTAFTQADSRLMDVGTLEPADSTIN-----NLVIN--- 495
DB 1237 ISGCTVEVK-----ATESLTTQNSKIKATGTGEANTSAFTIGCTISGTVVWTANAGD 1291
QY 496 -----ISSIDGA-----KRAKIETKATSKNLTLSGTTITLDPGTGYENHSLRNPOS 542
DB 1292 LVINGCAEINAEGAATLTTSGLKLTTEASSHTTSKAGOVNLSAODGSVAGSINAN-- 1348
QY 543 YDLELKASGVTS 556
DB 1349 ---VTLNTTGTTLT 1359

RESULT 7
US-08-302-832-2
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US Pct/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813

Oy	20	MAREYVLDSSAFDGNKNGNEVRSSQDACTYLLPFGNVTLEINPEOTGAITRSCFPN	79
Db	818	MIGKGIVAKKNITFEEGG-NITEGSRRKAVE-----IGCNVTINV--NANVTLLIGSDFN	868
Oy	80	TGKDLT-----FTGNGNSLIFQVVDAGTVAG-AAVNSSVDSKTFFIGESSLSF	127
Db	869	HOKPLTIKKDVIIINSGLNLAGNV-----NIAGMLITYESNANKRAIINFNFNWGL	920
Oy	128	IASEPG-SSITFGKAVSC-----STGSLSLTKNVS---LFPSKFSTDNGAIIAKTLJSL	178
Db	921	FDMKNGNISIAKGARFKDIDMSKLSTTNSSSYRTIISGNTKNG-----DLNI	974
Oy	179	T-T-GTYMSALFSENTSSKKGAIQTSIDLTT-----GNQGEYSRSDNTSSDSGAIF	229
Db	975	TNESDSTEMOIGGDVYQCKGNLTISSDKINTIKQITIKAVDGENSSODATNNANLIKT	1034
Oy	230	TEASVT-----IS--NNAKVSFDNKVTGASSSTGDSMGACAICAKTDPDKVTLTGNO	283
Db	1035	KELKLTODLINSGFNKAETIADKDSLTIIGNINSADGNNAKVTFPNQVKDSKIADSHKY	1094
Oy	284	LIFS-----NNTSTTAGAIIYVKKLELASGGLTLESRN-SVNGGTAPKGAIAEDS	334
Db	1095	TLHSKEVETSGSNNTDESSDN-----NAGLTIDAKWTVNWN-----ITSH	1135
Oy	335	GELSLSADSDIYF-LGNVTYSTPR-----TNMSSIDLG----TSAKMTALRSAGRVIYF	386
Db	1136	KAVISAITSEITTKTGTIMATTGWELLADOTGSILOGIESSGSVYTLPALEGALAV-	1193
Oy	387	YDPTTSSSTTVLVKNETPADSALOYTGGIIPTGKSLETEAAASKNLITSKLPVT	446
Db	1194	--SNISNIVTVANSGALLTTAGSTIGTESV-----TTSOSGDIGC-----T	1236
Oy	447	LSGGLSLKHGVLTQFAQTQOADSRLXEMDYOTLEPADISTIN-----NLEVIN---	495
Db	1237	ISGTVYEVK-----AESLITSONSKIRATTCGEANVTATGTCIGTISGNTVANTACD	1291
Oy	496	-----ISSIDGA-----KKAETRKATSKNLTLSGTLITLDPDTGYEHNLSLRPOS	542
Db	1292	LYVNGAEINATGCAATYLLTSSGKITLKTEKSSHITFAKOVMLSQDSSVAGSINAN---	1348

OY 543 YDILELKASGTVS 556
 Db 1349 ---VTLNTGTTLT 1359

RESULT 9

US-08-469-880-2
 Sequence 2, Application US/08469880
 Patent No. 5876733
 GENERAL INFORMATION:
 APPLICANT: Barenkamp, Stephen J.
 TITLE OF INVENTION: High Molecular Weight Surface Proteins
 TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Shoemaker and Matlare, Ltd.
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202-0286
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/469,880
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9205704.1
 FILING DATE: 16-MAR-1992
 PRIOR APPLICATION NUMBER:
 APPLICATION NUMBER: US PCT/US93/02166
 FILING DATE: 16-MAR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/302,832
 FILING DATE: 16-SEP-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Berkstresser, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1536 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-469-880-2

Query Match 5.08; Score 240.5; DB 2; Length 1536;

Best Local Similarity 22.68; Pred. No. 5.3e-10;

Matches 139; Conservative 106; Mismatches 220; Indels 149; Gaps 29;

OY 20 MATEYVLDSSAFDGNKNGFVSRESOEDAGTYLFEKGNVLTLENIGTGTAKRSCFNN 79
 Db 818 MIGGIVAKKNITPEEG-NITFGSRKAYTE-----IEGNTINN--NANTLTGSGFDN 868
 OY 80 TKGDLT-----FTGNGSLLFOTVDAGTVAG-AAVNSSVVDKSTTFIFSSLSF 127
 Db 869 HOKPLTIKKVLIINSGLNLTAGNIV-----NIAGNLVYESNANKATINFEFNGGL 920
 OY 128 IASPG-SSTITGKAVASC---STGSLSLTRKNS---LFSKRFSTDNGAIAKTLSL 178
 Db 921 FDNKGNISISIAKGARFKIDNKNLSITTNSSSTYRTIISGNITNKNK-----DLNI 974

OY 179 T--GTTNSALFSENTSSKKGAIQTSDALTT-----GNGEVSFSDNTSSDSCGAIF 229
 Db 975 TNEGSDTEMQIGDVSOKEGNLTISSDKINITKQITIKAGVDENSDATNNANLTIKT 1034
 OY 230 TEASVT-----IS--NNAKVFIDNKVTGASSSTTGDMSGAICAKTSTPTKYLTLGNOM 283
 Db 1035 KELKLTODLNLISGFKNKKEITAKDSDTLTIGNTNSADGTNAKKVTFNOVKRSKISADGHV 1094
 OY 284 LIFS-----NNTSTTAGAIYVKKLELASGGLTFESRN-SVNGSTAPKGAIAEDS 334
 Db 1095 TLHSKVTSSGNNTTEDSSN-----NAGLTIDANKVYNNN-----ITSH 1135
 OY 335 GEHSLADSGDIYF-IGNVTSTTPG---TNRSSIDG---TSAKYTLRSAAGRAIYF 386
 Db 1136 KAVISATSGEITTKGTGTINATGVNEITAGTISLGGIESSSGSVTLTATGALAV-- 1193
 OY 387 YDPITGSSSTTVDVLVKNTFPADSALOYGNIIIFGEKISETEADSKRLTKLQPYT 446
 Db 1194 --SNISGNTVTVANSALTTLLAGSTIKTESV-----TTSOSGDIGC-----T 1236
 OY 447 LSGGTLSEKHGVTLOQAFTOQADSRLMDVGTLEPADSTIN-----NLVIN--- 495
 Db 1237 ISGTVYEVK-----ATPSLTQSNKSIKATGGEANVTSACTIGTISGNTVNTANAGD 1291
 OY 496 -----ISSIDGA-----KKAIEFKASKNULTLSGTTLLDPTGTFYENHSLRNPOS 542
 Db 1292 LTVNGCAEINATGCAATLTTSKGLTTEASSHTTSAGQVNLASODSGVAGSINAAN-- 1348
 OY 543 YDILELKASGTVS 556
 Db 1349 ---VTLNTGTTLT 1359

RESULT 10

US-08-728-470-2
 Sequence 2, Application US/08728470
 Patent No. 5928651
 GENERAL INFORMATION:
 APPLICANT: Barenkamp, Stephen J
 TITLE OF INVENTION: High Molecular Weight Surface Proteins
 TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Shoemaker and Matlare, Ltd.
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202-0286
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/728,470
 FILING DATE: 424
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/302,832
 FILING DATE: 16-MAR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US PCT/US93/02166
 FILING DATE: 16-MAR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9205704.1
 FILING DATE: 16-MAR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Berkstresser, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-633

```
? TELECOMMUNICATION INFORMATION  
? TELEPHONE: (703) 415-0810  
? TELEFAX: (703) 415-0813  
? INFORMATION FOR SEQ ID NO.: 2:  
? SOURCE CHARACTERISTICS:  
? LENGTH: 1536 amino acids  
? TYPE: amino acid  
? STRANDEDNESS: single  
? TOPOLOGY: linear  
? MOLECULE TYPE:  
? US-08-728-470-2
```

Query Match	5.0%	Score 240.5	DB 2	Length 1536
Best Local Similarity	22.6%	Pred No. 5.3e-10		
Matches 139, Conservative 106, Mismatches 220,				Indels 149, Gaps 29

QY 20 MIETVELDSSAPDGNKKNFVSREOEDAGTTFYFKGNVLEINPGTMTATSCFN 79
 Db 818 MIGKGIYAKKNITFEQG-NITFGSRKAYE-----IEGNVITNN--NANVTLLISDFPN 868
 QY 80 TKGDLT-----FTNGNSLLEFQTVDAQTAVG-AAVNSVYDKSTFEIGFSLSIF 127
 Db 869 HOKRLITKKDVIINSGLNITAGNIV-----NIGNLTVESNAFKAITNFTFNWGL 920
 QY 128 IASPG-SITITGGKAVSC---STGSLILTKNVS---LLFKNFTDNGAITKTKTSL 178
 Db 921 FDNKGNNSITAKGAFAFKIIDNSKNIUSTITNSSSTYRITISGNTINKKG-----DLWT 974
 QY 179 T-GTTMSALFSENTSSRKGAIGTSDALTIT-----GNQGEVSFSDNTSSDGAIF 229
 Db 975 TNESSDPEMOIGGIVSQKEENLTISSDKINITKQITIKAGVGDENSDPATNNAMLTIKT 1034
 QY 230 TEASVT-----IS--NNAKVSFIDNKVTGASSSTTGMSGGALCAVYITSDTFTVLLGNOM 283
 Db 1035 KELKLTIDDLNISCPNKAKEITAKGSDLTIGNTISADGTNAKKVTEIQQVYKDSKTSADGHV 1094
 QY 284 LLEF-----NNTSTTAGAIVYKRLKELASGLTLFERN-SVNGTAPKGAIAIEDS 334
 Db 1095 TLHSEFTSGSNNTEDSDPN-----NAGLTIDAKNVTYVNN-----ITSH 1139
 QY 335 GELSLSDSGSDIYF-LGNVTYSTTPG---TNRSIDLG---TSAMKTLRBAAGRIYF 386
 Db 1136 KAVSISTSEIITTKTIGTTINATGNVEITIAQCGSLTGGIESSGCVTLTATEGALAV-- 1195
 QY 387 YDPITTGSTTVVDLVKNETPAPDSALOYTGNIITFEKLESEHADSKNLTSKLLPOVT 446
 Db 1194 --SNISGNTVTVANSGLATLTLAGSTRIKGHEV-----TTSQSGDIDG-----T 1236
 QY 447 LSGGTLSEKRGVILQIQAFQOQADSRLEMDVGTTELPADTSIIN-----NLVIN-- 495
 Db 1237 ISGGTVEVK-----ATESLTTOSNSKIKATGTGANVTSATIGTIGTISGNTVNVANMGD 1291
 QY 496 -----ISIDGA-----KKAKEITKATSKNLTLGSGITTLDPRTGFPEHNSLRNPOS 542
 Db 1292 LTVNGAENIATGGAATLTTSSKRLTTEASHTTSKAGOVNLSADGSAVGSINAN--- 1348
 QY 543 YDLIELKASGTVS 556
 Db 1349 ---VTLNLTGTLTT 1359
 RESULT 11
 US-08-617-697-2
 ; Sequence 2, Application US/08617697
 ; Patent No. 5977336
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Maltare, Ltd.

```

1 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
2 STREET: Bldg. 1
3 CITY: Arlington
4 STATE: Virginia
5 COUNTRY: U.S.A.
6 ZIP: 22202-0286
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: floppy disk
9 OPERATING SYSTEM: IBM PC compatible
10 SOFTWARE: Patentin Release #1.0, Version #1.30
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: US/08/617,697
13 FILING DATE: 01-APR-1996
14 CLASSIFICATION: 424
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER: US 08/302,832
17 FILING DATE: 05-OCT-1994
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: US PCR/US93/02166
20 FILING DATE: 16-MAR-1993
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Berkstresser, Jerry W
23 REGISTRATION NUMBER: 22,651
24 REFERENCE/DOCKET NUMBER: 1038-557
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: (703) 415-0810
27 TELEFAX: (703) 415-0813
28 INFORMATION FOR SEQ ID NO: 2:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 1536 amino acids
31 TYPE: amino acid
32 STRANDEDNESS: single
33 TOPOLOGY: linear
34 MOLECULE TYPE: DNA (genomic)
35 US-08-617-697-2
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
96
```

```

QY 447 LSGGTLKKGCVTLQTAFTQOADSRLMDVTLLEPADSTIN-----NLVIN--- 495
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1237 ISGTVLEVK-----ATESLFTQSNKSIKATGEANVTSAITGCTIGTANTVTAAGD 1291
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 496 -----ISSIDGA-----KKAKIEKATSKNLTLSGTLTLPDPTGTFVENSRLRPOS 542
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1292 LTVNGAEINTEGAALITTSKGLTTEASHITTSANGOVNLSAQDSVAGSIYAAAN--- 1348
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 543 YDILEKASGVTS 556
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1349 ---VTLNWTGTLTT 1359
| : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 12
US-08-194-290-7
; Sequence 7, Application US/08194290
; Patent No. 5500353
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; TITLE OF INVENTION: Bacterial surface protein expression
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shlesinger, Arkwright & Garvey
; STREET: 3000 South Eads Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garvey, George A
; REGISTRATION NUMBER: 17737
; REFERENCE/DOCKET NUMBER: 5946-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-5600
; TELEFAX: 703-836-5288
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1026 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-194-290-7

Query Match 5.0%; Score 238.5; DB 1; Length 1026;
Best Local Similarity 24.1%; Pred. No. 4e-10;
Matches 178; Conservative 86; Mismatches 258; Indels 217; Gaps 33;

QY 12 TFAFTPLSMATEYVLDSSASFDKNKNGNSYRESQEDAGTTFYFKGNVLEINP-----G 67
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 269 TEVAGEVAGAAATLVGDTLGGAGT-----DVLNMYQAAAVTALPTG-VYISGETINVT 322
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 68 TGAFTKSCFNNTKGDLTFTGNGNSLLFOYVDAG-----TVGAAVNSVSVKSTFEI 120
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 323 SGAAITLNTSSGVTG-LTALNTNLSGAAQYVTAAGONLTITTAQAQANNAVAVD----- 375
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 121 GFSSLSFTASPGSSITTGKGAVSGSTGSLFTKNVSLFSKNFSIDNGALITATLSLTG 180
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 376 GGAVTVASGTGVTGTTVGANSAAAGTAVS-----SVANSSTTTGAIATVG 423
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 181 TTKMALFSENTSSKKGAIGTQSDALTTTGNG--EVSFSDMTSSDSGAAT--TEASVTI 236
| : : : : : | : : : : : | : : : : : | : : : : : |

```

```

Db 424 GT--AVTVAOTAGNAVNTITQADVTVTNGSSTTAVTQTOTAATAGATVAGVNGAVTI 481
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 237 SNNA-----KVSEIDNKVTGA-----SSSTGDMG-----GAICAYKSTDT 274
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 482 TDSAAAGATTAGKATVATVILGFCGAATIDSSALFTVNLGCTGSLGIGRALT----- 524
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 275 KVLITGOMLFFSNNTSTTGAGAIYKKLELASGGLTFERSNVNGSTAPKG-GAIAIED 333
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 535 --TPTANTLTILNGLTIT--GAITDSEAAADG---FTTINAGSTASTIASLVAAD 586
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 334 SGEISLSDSGDIY-----FLGNVYTSF-----TPGTRRSSIDLGTS 371
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 587 ATTINISGDARVITTSHTAALGITVTNMGVATGLGELATGLVTCGACADSLILGAT 646
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 372 KMTALRSAGRAIYFDPITGSSITV-----TDVLKN----- 405
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 647 K--AIVMGAG-----DDTVTSATLIGAGSVNGCGDGTDLVAVNNGSSFSAPAFGGF 698
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 406 ET--PADSALOYGNII--FVGEKLSFEPA----- 432
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 699 ETLRVAGAAAGSHNANGFALDLAGATTTTNNVAVNGVLVLAPTGTTVTLANAT 758
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 433 ---DSKNLTSKLLQPVTLSSGTLSL-----KRGVTLQTA-----FT 466
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 759 GTSDFVFNLT--LSSAALAGTVALAGVETVNIATDTNTTAHVDFLLOATSAKSIVT 816
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 467 QOADSRLMDVGTLEPADSTI-----NNLVINISSIDGAKAKIETATSK 514
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 817 GNGAGNLNTGNTAVTSFSDASAVTGSAVTEVSNLTVEVVTIRGACADSLTGSATA 876
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 515 NLTL---SGTITLDPDTGTFVENSRLNPOSYDILELKASGTVTSAVTPDPIMGEKFHY 571
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 877 NDILIGAGADTLVYIGT-----DTFTGIGADIIFDINALIGTSTAFVITIDAAGDKLDL 932
| : : : : : | : : : : : | : : : : : | : : : : : |
QY 572 GYGCTWGPYWG--GASTT 589
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 933 VGISTNGAIDAGFAAVT 951
| : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 13
US-08-169-927-2
; Sequence 2, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Wei
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/742,128
; FILING DATE: 08/09/91
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, A. David

```



```

Oy 277 TLTNOMLFSN--TSTTAC-----AIYKTEL-----ASGGLT 311
Db 1091 SADOHNATLMSKYKVTSSNGCRSNDNDGTGLTITKANVEYKDDITSLKTIVTASEKVT 1150
Oy 312 LFSNSVNGTAPKGAIAIEDSGELSLSDGDIVLGNTVSTTPGTNRSSIDL-- 368
Db 1151 TTACSTT--ATNGKASITTKGDIS--GTSIGNTSVAETADLTTRK$SK$EASGEAN 1206
Oy 369 -TSAKMTALSAACARAIYFDPITTSSTFVPEVLVNEFTPADSLQYGNIIFF--CEK 425
Db 1207 VTSATGTIGGISNIV---NVTANAGDLTVYGAEINMTGCAATITATGNLTITTEADSS 1266
Oy 426 LSETEAADSKNLTSKLLQPYTLISGGLTSLKHGYTLIOTO-AFTQOADSRLMDVGTLEPA 484
Db 1264 ITST-----KGVDLLAQNNGSINGSIWAA--NVTLNTGTGLTFTVAGSDJKAATSGT----- 1311
Oy 485 DSTINNUVINISSIDAKKAKKI-----EKAATSKNLTLSGTTI 523
Db 1312 -----LVIN-----AKADKALMGASGDSTEVNANVNASGSGSYT 1344

```

RESULT 15
 US-08-302-832-4
 Sequence 4, Application US/08302832
 Patent No. 5603938
 GENERAL INFORMATION:
 APPLICANT: Barenkamp, Stephen J
 TITLE OF INVENTION: High Molecular Weight Surface Proteins
 TITLE OF INVENTION: of No. 5603938-Typeable Haemophillus
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Shoemaker and Mattare, Ltd
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 STREET: Bldg. 1
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202-0286
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/302,832
 FILING DATE: 16-SEP-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9205704.1
 FILING DATE: 16-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US pct/us93/02166
 FILING DATE: 16-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Berkstresser, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1058-404
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1477 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

Query Match	4.7%;	Score 223.5;	DB 1;	Length 1477;
Best Local Similarity	23.5%;	Pred. No. 1e-08;		
Matches 138;	Conservative 93;	Mismatches 212;	Indels 143;	Gaps 29

Oy	38	NGNSVSEOSD-----ACTY-----LFGNVTLENIPEGTGAIRKSCNNPTKGLT	85
		: : : : : : : : : :	
Dd	802	NSNFSLKQIDDFIDGYARAIAINSTINISILGCVNTLGG-QNSSSTIGNITIEKAANYT	860
Oy	86	FTGNG-----NSLFFQTVBDGATVAGAAVNSSV-----VDKSTFIEGSSLS	126
		: : : : : : : : :	
Dd	861	LEANNAPNOQINDRVIKGLSLLVN-----GSLSLGEMNDIGNLITSSAIFPK-KTRD	915
Oy	127	FIASPGSITTGKAGVSCSTGSLSTLK-----NVSLFKNFSTDNCGAITYAK--TLS	177
		: : : : : : : : : : : : :	
Dd	916	TLNTGTNFTNGTAIEINITGGVVKLQNVNDGDNLNITTHAKRNOBSIIGDIIKNGSLN	975
Oy	178	LTGTMSA--LFSNNTSKKGAIOGSDALLT-----GNQGVFSDNTSSDSCGAI	228
		: : : : : : : : : : : : :	
Dd	976	ITDSNDNAEIOIGCNISQKGNLTISSDKINITKQITIKKGIDGEDSSDAR-SNANLTI	1034
Oy	229	FT-----EASVATISNNNAVSFIDNRKVTCASSSTTGDMSGAICAYKTSY-----DTRV	276
		: : : : : : : : : : : : :	
Dd	1035	KTRKELKLTEDLSISGFPAKAITAKDGRDLTIGSNONGSGAIA-----KTVFFNNVKDSKI	1090
Oy	277	TLTGNOMLFESN-NTSTTAG-----AIYVKLEL-----ASGLT	311
		: : : : : : : : : : : :	
Dd	1091	SADHNVTILNSKYVSTSSNGCRESNDNDTGTLITRAKNEVKKDITLSKTIVITASEKVT	1150
Oy	312	LFSNNSVNGGAPRGGAIAIEDGSELSSADSDIYFLGNVTYTTTPGNNBSIDLQ-----	368
		: : : : : : : : : : : :	
Dd	1151	TTYASSTN--ATNGKASITTKTGDIS-CTISQNVVSATYDLTKTSQSKIEAKSGAN	1206
Oy	369	-TSAKMTALRSAGRAIFYDPTTGTSSSTTVTDVLKVNFTPADSALOYTGNIIFP--GEK	425
		: : : : : : : : : : : :	
Dd	1207	VTSATGIIIGGISNY--NTANAGDLTVNGAELMTBEAAILLATGNTLTTEAGSS	1263
Oy	426	LSFTEADSKNLTSKLLQPVYLSGGTSLKHGVTLOTO-AFTQOADSRLMDVGTTLPEA	484
		: : : : : : : : : : : :	
Dd	1264	ITST-----KGVDLLAONSGISINGINAA--NVTLWTTGTLTTVAGSDIKATSGT-----	1311
Oy	485	DTSTINNLVINISSIDGAKKAKI-----EIKATSKNLTLGCTIT	523
		: : : : : : : :	
Dd	1312	-----LVIN-----AKAKALINGDASGDSTEVNANVNSGGSGSVT	1344

Search completed: November 20, 2000, 11:20:03
Job time: 10857 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2000, 09:30:31 ; Search time 14.34 Seconds
(without alignments)
4106.770 Million cell updates/sec

Title: US-09-428-122-2

Perfect score: 4774
Sequence: 1 MKSSFPKFEVSTFAIFPLSM.....FELRGSSRNRYNDLGAKYQF 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR-65:*

1: pirl:*
2: pirl:*
3: pirl:*
4: pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2058	43.1	928	2 G81591	polymorphic membra
2	1982	41.5	928	2 B72077	polymorphic membra
3	1965	41.2	936	2 C72078	polymorphic outer
4	1964	41.1	936	2 B81591	polymorphic membra
5	1940	40.6	930	2 A81591	polymorphic membra
6	1936	40.6	930	2 D72078	polymorphic outer
7	1855	38.9	928	2 D72077	polymorphic outer
8	1855	38.9	949	2 F81591	polymorphic membra
9	1573	32.9	841	2 E72130	polymorphic membra
10	1444	30.2	1276	2 G81591	polymorphic membra
11	1442.5	30.2	922	2 B72131	polymorphic outer
12	1441.5	30.2	922	2 F81539	polymorphic membra
13	1383.5	29.0	1407	2 B72078	polymorphic outer
14	1377.5	28.9	973	2 F72076	polymorphic outer
15	1377.5	28.9	995	2 C81593	polymorphic membra
16	120.5	23.5	1013	2 G71460	probable outer mem
17	1051	22.0	987	2 H81722	polymorphic membra
18	911	19.1	867	2 F81721	polymorphic membra
19	863	18.1	878	2 B71460	probable outer mem
20	685.5	14.4	1609	2 H72013	polymorphic membra
21	672.5	14.1	978	2 G72076	polymorphic outer
22	667.5	14.0	978	2 B81593	polymorphic membra
23	666	14.0	947	2 D72067	polymorphic membra
24	666	14.0	1723	2 E72067	polymorphic membra
25	665.5	13.9	1732	2 C81601	polymorphic membra
26	665.5	13.9	946	2 D81594	polymorphic membra
27	665.5	13.9	946	2 C72075	polymorphic outer
28	647	13.6	952	2 D81593	polymorphic membra
29	646	13.5	934	2 A72075	polymorphic outer

30	603.5	12.6	938	2 H72074	polymorphic membra
31	603	12.6	975	2 F71518	hypothetical prote
32	597	12.5	1016	2 H71460	probable outer mem
33	594	12.4	983	2 A81723	polymorphic membra
34	587.5	12.3	986	2 B81675	polymorphic membra
35	576.5	12.1	1531	2 H71468	probable outer mem
36	559	11.7	1751	2 G71518	hypothetical prote
37	549	11.5	976	2 F81722	polymorphic membra
38	548.5	11.5	1672	2 C81675	polymorphic membra
39	544	11.4	1520	2 A81731	polymorphic membra
40	541.5	11.3	964	2 E71460	probable outer mem
41	533.5	11.2	1450	2 D81675	polymorphic membra
42	523	11.0	976	2 A71517	hypothetical prote
43	490	10.3	1034	2 F71460	probable outer mem
44	475.5	10.0	514	2 E72076	polymorphic membra
45	453	9.5	1025	2 G81722	polymorphic membra

ALIGNMENTS

RESULT 1

G81591

polymorphic membrane protein G family CP0303 (imported) - Chlamydomonas reinhardtii (s)

C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii

C>Date: 31-Mar-2000 #sequence: revision 31-Mar-2000 #text: change 11-May-2000

C/Accession: G81591

R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydomonas reinhardtii trachomatis Mopn and Chlamydomonas reinhardtii AR39

A:Reference number: A81500, MIMD:20150255

A:Accession: G81591

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-928 <REA>

A:Cross-references: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AAF38160.1; PID:g718

A:Experimental source: strain AR39, HL cells

A:Genetics: CP0303

Query Match	Score	43.1%	Length	928
Best Local Similarity	46.7%			
Matches 444; Conservative 142; Mismatches 320; Indels 44; Gaps 20;				
QY 1 MKSSFPKFEV-STFAIF-PLSMATEVLDSASFQGNK-GNFSVRESQEDAGTTL 55				
DB 1 MKSFSMLVLSSTLACTSCSTVEAATNAENIGPSDFSGTGTVPKNT--TTGIDY 58				
QY 56 FKGNVLENIPIGCTATKSCFNNTKGLDFTGNGSNLLQYDAGVAGAAANSSVVDK 115				
DB 59 LTGDTLQNL-GDGAALTKGFSFDTESLSPAGKGLSLFLNLS-SAEAAAL-SVYTDK 115				
QY 116 STTFGIISSLFASPGSSITTT--GKGAVCSSTGSLTFLNLSLFSKNTSTNGCAITA 173				
DB 116 NLSTGSSSLFLAAPSVAITTPSGKAVKCG-GDLTFDNNGLITFLFQDCEENGCAIST 174				
QY 174 KTLSTGTTMSALPSENTSS--KKGAIOTSDALTTGNGEVSFSDNTSSDGAATFT 220				
DB 175 KNLSTKSTGSISEFGKSSATGKKGAIATGIVDTNTATFLFSNNIAEAGCAINS 234				
QY 231 EASVTISNNNAKVSFIDNKVTCGASSSTGDSGGAICAYKTSPTKVTLTGNOMLFSNNT 290				
DB 235 TGNCTITGNTSLVFSNSVT---ATAG--NGAHL-----SGDADVTISGNSVTSFGNQ 283				
QY 291 STTAGAIYVKKLELAS--GGLTLFSRNSVNGGTAPKGAIAIEDSGELSLSDSGDIV 348				
DB 284 AVANGCAIYAKKLTLAGSGGGGGSISNNIVQGTAGNGAISILAGECSLSEAGDITF 343				
QY 349 LGNTVITTP-GTNRSSIDLTGSKAKTALSAAGRAIYFDPITTGSTIVTVLKVNET 407				
DB 344 NGNAIVATTPQTKRNSIDIGSTAKITLRAISGHSIFDFYDPTANTADSTDTLNLKA 407				

QY 408 PADS:LQYTGIIITGKLESTFEADSKNLTUSKLLQPYTLPSGTLTKHCVTLQIQATQ 467
Db 404 DAGNSIDYSGSIVISGEKLSDEAKVADNLVSTLQPYTLTGNLVLKRGVTLDTGTGTO 463
QY 468 QADSRLMEMDVTTLPEPA-DTSTINNVLVINISSIDAKKAKIETKATSKNLTLSGTLTLD 526
Db 464 TAGSSVIMADGATTLKASTEEVTLTGLSIPVDSLGSGKKVVAASASKNVALLSGPILLD 523
QY 527 PGTGFYFNHSLRNQSDVILDLKASGIVTSTAATPDPINGEKFHYGTQGWPIW---- 562
Db 524 NQGAENHDLGKTQDSFVQLSALGATTTDDVPVPAVPAIPHXYGQFWG-MTWVDDT 582
QY 583 -GTGASTATTEMTKTGYIPNPERIGSLVPNSLMAAFDISSLHYMETANEGLQGRAF 641
Db 583 ASTPKTATLTAIWNITGLPLNPEROGPLVPNSLMGFSFDIQAGVTESSALTLCSDRGF 642
QY 642 WCAGLSNFHHKDSFKTRRRGPHHLSGGYVIGNLHTCSDKILSAFCQLFGRRDHYVAKN 701
Db 643 MAAGVAAFLLDKDKGKKRKYRHKSGVYAGAAQCSSENLISFAFCQLFGSDKFLVAKN 702
QY 702 QGTVYGGTLYVQHNETYISLCKLRPCSLVYP---TELPLFSGSLSTYHNDNLKTKY 758
Db 703 HTDTYACAFPIQH---ITECSGFTGCLDKLPGSWSHKPLVEGOLAVSHVNDLKTKY 758
QY 759 TTYPTVYSGWGNDSFALEFGGRAPICLDESALEQYMPMKIQFYVAHGEKEQGTTEAR 818
Db 759 TAYPEYKSGMGNNAFMNMLGASSHSPYELHCFQDYAPAIKTLNLYIRDSFEKTEGR 818
QY 819 EFGSSRLVNLALPGIRFDKESDCODATYNTLTGYTVDLVRSNPDCTTLRLISGDSWKTF 878
Db 819 SFDDSNLFENSLPLPGVFKFEKSDCNDFFSYDILLTSVPDLIRNDPKCTTLYVIGSASMETY 878
QY 879 GTNLARQALVLRAGNHFCFNSNFENAFSQSFELRGSSRRNYNDLGAQYOF 928
Db 879 ANNLRQALQVVRAGSHAFSPMFEVLGQFVEVRSSSRINYNDLGRKPEP 928

RESULT 2
B72077
polymorphic membrane protein G family CP0306 [imported] - Chlamydia pneumoniae (stra
C.Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C.Accession: B72077, B81592
R.Katman, S.; Mitchell, W.; Marathe, R.; Lamnel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A.Reference number: A72000; MUID:99206606
A.Accession: B72077
A.Molecule type: DNA
A.Residues: 1-928 <RNR>
A.Cross-references: GB:AE001628; GB:AE001363; NID:g4376730; PIDN:AA018591.1; PID:g437673
A.Experimental source: strain CWL029
R.; Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gilm, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A>Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A.Reference number: A81500; MUID:20150255
A.Accession: B81592
A.Molecule type: DNA
A.Residues: 1-928 <REA>
A.Cross-references: GB:AE002192; GB:AE002161; NID:g7189226; PIDN:AAF38163.1; PID:g718923
C:Genetics:
A:Gene: pmp_9; CP0306

	Query Match	41.5%	Score 1982;	DB 2;	Length 928;
	Best Local Similarity	45.2%;	Pred. No. 6.6e+100;		
	Matches 429;	Conservative 150;	Mismatches 327;	Indels 44;	Gaps 19;
Oy	1 MKSSPPKVFSTFAIFPLSM-----IATETVLDDSSASFDGKNKGNFVSYSQEDAA-GTT	53			
	: :	:	:	:	

```

Db      1 MKSLSLHWPILSSLSLAPLSLNSFAAAYEINLGPNTSFG--PCTYTPPAQTTTA
Qy      54 YLEKGNVTLLENIPGTGATITKSCFNNTKGDLTFTNGNSLLFQIVDACTVAGAAYN
Db      59 YNLIDVDSFTN-AGSPRLTALASCPEKETTGNLSFQHGQFPLONIDAG-ANCTFTN
Qy      114 DKSTFIIGSSLSFIASPPSSITTTGKAVSGTGSLSLTNNVLSLFKNSFNKDNGALTA
Db      116 NKLLSFSGSYLSLIQT--TNATITGAIK-STGACSIQSNVSCFEGONFSDNGALOG
Qy      174 KLLSLGTGTMSALFSEBNTSSKKGAIGTSDALTITGNQGEVFSFSDNTSSDSGALFTEAS
Db      173 SSISLIS-LMPNLTLPKKNATOKGALYSTGCTITNTLNTINSASFENMTAANNGAITYEAS
Qy      234 VTISSNAKVSFLDNKVATGASSTTGDMSGAIICAKTSTDRKV-LLTGNOMLTSNNST
Db      232 SFTSSNKALSFNNSTYATSA-----TGGALYCSTAPKPVLTLSNGELNLTGNAT
Qy      293 TAGGAIVYKKELELAGGLTLPFSRNSVNGTAPKGAIAIEBGSGLTSADSGDIYELNT
Db      286 TGGGAIYTDNLVSLSSGPTLPFKNNSAIDTAPALGGAIAADSGSLSLALGDITFEGNT
Qy      353 V---TSTTPGTRRSIDG-TSAKMTALRSAGRAIYHYDDITTCSSITTVDYLVKNTP
Db      346 VYKGAASSQTTTRNSINIGNTNAKIVQLRAQGNITFYDDITSTIALSDALNLPD
Qy      409 ADSALQYTGNIIFTGKLESETAEADSKNLTKSLQPLVLTSGTSLKAGVILQFOAFQ
Db      406 IAGNPAYQGTIFYSEGEKISEAAEAADNLKSPIQPLTLAAGQSLKSGVILVAKSFQS
Qy      469 ADSRLMDVGTTLERADTSTINNVYINISSIDGAKAKAIEKRAISKNLTLSTGTTILDPT
Db      466 POSTLIMAGTTLERADITINNVLNVDISKETKALKATQASQVYTLGSLSLVDP
Qy      529 GTFFENHSLRNPQSDIIELKAS--GTVTSAVTPDPIMGEKFFHYGOGTGPYWGCA
Db      526 GNVEYEDVSWNNQVPSCLITLADDPANHIDDIADPLEKPNIMHYOGNMA-LSMQEDT
Qy      587 ST---TAFNNTKTGYINPERIGSLVNSLMAAFIDISSLHYLMEFANBEOGDRAFC
Db      585 AKRSKAALTLWKGYNPNPERGVLANTLWGSFVDRSIOOLVATVRSOETRGKWC
Qy      644 AGLSNFFHKDSTKTPRGFRHLSGGVIGNLHTCSDKLTLSAFCOLFGDRDGYVAKNOG
Db      645 EGISNFFHKDSTKIKNGFRHISAGYVAGATTTLASDNLTITAFQOLREKQDHIKNKRA
Qy      704 TVYGGTILYYQHNEYIISLPCKLRPQSLYVP--TEIPVLFSGNLSTYHNDLTKRYTT
Db      705 SAYAASLHLOHATLSS-----PSLLRLYLGSESEQEPFLDAOISYISKNTKRYTQ
Qy      761 YPTVAGSGNDSPALFEGGRAP-ICLDSEALFEDYMPMKLOFYAAHQEGKEGTE-AR
Db      759 APKGSSSYNGOCALELASSLPHTALSHEGFLFAHYFPPIKEASTIHODSKENNTLVR
Qy      819 EFGSSRLYNLALPIGIRPDKESDQDATYNNLTIGTYDVLVRSNPDCTYTLRIISGDSMKT
Db      819 SPDSGDLINVSPIGTPFERESRNERASYEATYIVYADVVRKPNDCDTALLINNTSWKT
Qy      879 GTNLARQALVLRAGNHFCFNSNFEAFSOFSELAGSSSNYNVWDGAKYQF
Db      879 GTNLSRQIGIRAGIFFAFSPNLEVTNLSLMEIGSSSSYNVADGGKRFQF

```

RESULT 3
C72078
polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL0299)
C.Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C.Accession: C72078
R.Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
Nature Genet. 21, 385-389, 1999
A.Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
Reference number: A72000; MUID:99206506

A:Accession: C72078
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-936 <ARN>
 A:Cross-references: GB:AE001627; GB:AE001363; NID:g4376721; PIDN:AAD18589.1; PID:g437672
 A:Experimental source: strain CWL029
 C:Genetics:
 A:Gene: pmp_7

Query Match 41.2% Score 1965; DB 2; Length 936;
 Best Local Similarity 43.4%; Pred. No. 5 6e-99;
 Matches 411; Conservative 179; Mismatches 327; Indels 30; Gaps 16;

1 MKSFFKVFESTFAIF-PLSMIATEVLDSS-ASFDGNKNGNSVRSOE-DAGTYLTK 57
 1 MKSSVSWLFESSILPFLSSLSIVAAEVLDDSSNNNSYDGSNGTTFVFESTDAAGTYSL 60
 58 GNVLTENIPGTGTAITKSCFNNTKGDLTFTGNGNSLLFQYDADCTAGAAVNSVDKST 117
 61 SDVSFQNAAGALGIPLASGCTLEAGDGLTFQGNQHALKFAINMGSSAGTAASTADK 120
 118 TFIFFSSLSFIASPGSSIT-TGKGAVSCSTGSLTKVNSLFSKNEFTDNGAITAKTL 176
 121 LFNDSFRLSIISCPSSLSPGQCALK-SVGNLSLTGNSQIIFTQNFSSDNGVINTKNF 179
 177 SLVGTMSALFSENT--SSKKGAIQTSALTLITGNOGENVSFSDNTSSDGAITFEASV 234
 180 LLSGTSOFASFSTRQAFQKOGGVYATGTTIENSPIVFSQNLAKGSGALYSTDNC 239
 235 TISNNAVSEFIDNKVGTGASSSTTGDMSGATCAVKTSTDTKVTLTGNOMLFSNNTSTA 294
 240 SITDNFQVITDGNAMAWAQA-----OGGAICC--TTTDKVTILTGNKNSLFTNNALTY 292
 295 GGAIYVKKLELASGGLTFERSNVNGTAPK--GGAIAIEDSGELSLSDSGDITVFLGNT 352
 293 GGAIISGLKVSISAGPTLFQSN-ISGSSAGOGGGAINIASAGELALJATSGDTFNNQ 351
 353 VTSTPTGNSSIDLGTSAKMTALRSAGRAIYFDPITGSSSTTVYDVKVNETPADSA 412
 352 VTNGSTST-RNAINIIDTAKVTISRAAGOSIYFDPITNGTAASDTLNLADANSE 410
 413 LOYGNIIFTGEKLESEADSKNLSKLLDPTVLTSGTSLKKGVTLOTQAFQOADS 472
 411 IEVGALVFSEKSLPEKAIANAVSTIRQPAVLARGDLVLRGVYVTFEDLTQSGSR 470
 473 LENDVGTLEPADTS-TINNLVINISSIDGAKKAKIETKATSKNLTLSGTTLLDPGTG 531
 471 ILMGGTTLISAKKANLSLNLAVLSSLDGTNKALKTEADKNISLSTGTTALIDTEGSF 530
 532 YENHSLNPOSYDILEK--ASGTVSTAVTPDPIMGKHYGOGTWGPIYVGTGAST 588
 531 YENHNLKSASTYPLLELTITGANGTITLGAISTLTLOEPETHYGQGW-QLSMANATSS 589
 589 -TATFNWTKGTGYPNPERIGSLVNSLMAFIDISSLHYMETANEGLOGRAFWCAGLS 647
 590 KISGINMTGRGYIPSPERKSNLPLNSLMGNFIDIRSNOLLETSSGEPERELMISGIA 649
 648 NEFHKDSIKTRRGFRHLSCGVIVICGNLHSCDKLTLAFCOLEGRDPIYVAKNOGTVYG 707
 650 NEFVRDMPTRHGRHISGTYALGITATTPAEDDLTFAFCOLFARDNNHITLKHNGDTYG 709
 708 GTLYOHNEHYISLPCLT-----RPSLSVYPTIEIYVLPSGNSLSTHTDNDLTKXTTY 761
 710 ASLTFHTTEGLFDIANFLWGRATRAPVLSLSQIIPSEPAKSYLHTDHMTYTTDN 769
 762 PYVSGWNSDFALEFGRAPICHDESALFEQYMPFMKLOFVYAHOGGEFEAREFG 821
 770 SIIGSMRNDAFCADLASLPFVIVSVYLLKEVEPVKQVYIAHQDPERNAHEGAFN 829
 822 SSRVLNALPLGIFEDKESDQATYNTLGLTYDVLVRSNDCPTTTRISGDSKKTGTN 881
 830 KSELINWEIPIGVFERDSKSEKGTDLTLMYILDAYRRNPKCOTSLIASDANMAYGTN 889

QY 882 LARGALVLRAGNHCFNFSNEAFSQSFELRGSSRNRYNDLGAAYOF 928
 DB 890 LARGGVSVRANHFOVNPHEIFQGFAYEVRSRRNYNTNLGSMFCF 926

RESULT 4

polymorphic membrane protein G family CP0308 [imported] - Chlamydia pneumoniae (s
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: B81591
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salze
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
 A:Reference number: AB1500; MUID:20150255
 A:Accession: B81591
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-936 <REA>
 A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AAF38165.1; PID:g718
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0308

Query Match 41.1% Score 1964; DB 2; Length 936;
 Best Local Similarity 43.4%; Pred. No. 6.3e-99;
 Matches 411; Conservative 179; Mismatches 327; Indels 30; Gaps 16;

1 MKSFFKVFESTFAIF-PLSMIATEVLDSS-ASFDGNKNGNSVRSOE-DAGTYLTK 57
 1 MKSSVSWLFESSILPFLSSLSIVAAEVLDDSSNNNSYDGSNGTTFVFESTDAAGTYSL 60
 58 GNVLTENIPGTGTAITKSCFNNTKGDLTFTGNGNSLLFQYDADCTAGAAVNSVDKST 117
 61 SDVSFQNAAGALGIPLASGCTLEAGDGLTFQGNQHALKFAINMGSSAGTAASTADK 120
 118 TFIFFSSLSFIASPGSSIT-TGKGAVSCSTGSLTKVNSLFSKNEFTDNGAITAKTL 176
 121 LFNDSFRLSIISCPSSLSPGQCALK-SVGNLSLTGNSQIIFTQNFSSDNGVINTKNF 179
 177 SLVGTMSALFSENT--SSKKGAIQTSALTLITGNOGENVSFSDNTSSDGAITFEASV 234
 180 LLSGTSOFASFSTRQAFQKOGGVYATGTTIENSPIVFSQNLAKGSGALYSTDNC 239
 235 TISNNAVSEFIDNKVGTGASSSTTGDMSGATCAVKTSTDTKVTLTGNOMLFSNNTSTA 294
 240 SITDNFQVITDGNAMAWAQA-----OGGAICC--TTTDKVTILTGNKNSLFTNNALTY 292
 295 GGAIYVKKLELASGGLTFERSNVNGTAPK--GGAIAIEDSGELSLSDSGDITVFLGNT 352
 293 GGAIISGLKVSISAGPTLFQSN-ISGSSAGOGGGAINIASAGELALJATSGDTFNNQ 351
 353 VTSTPTGNSSIDLGTSAKMTALRSAGRAIYFDPITGSSSTTVYDVKVNETPADSA 412
 352 VTNGSTST-RNAINIIDTAKVTISRAAGOSIYFDPITNGTAASDTLNLADANSE 410
 413 LOYGNIIFTGEKLESEADSKNLSKLLDPTVLTSGTSLKKGVTLOTQAFQOADS 472
 411 IEVGALVFSEKSLPEKAIANAVSTIRQPAVLARGDLVLRGVYVTFEDLTQSGSR 470
 473 LENDVGTLEPADTS-TINNLVINISSIDGAKKAKIETKATSKNLTLSGTTLLDPGTG 531
 471 ILMGGTTLISAKKANLSLNLAVLSSLDGTNKALKTEADKNISLSTGTTALIDTEGSF 530
 532 YENHSLNPOSYDILEK--ASGTVSTAVTPDPIMGKHYGOGTWGPIYVGTGAST 588
 531 YENHNLKSASTYPLLELTITGANGTITLGAISTLTLOEPETHYGQGW-QLSMANATSS 589
 589 -TATFNWTKGTGYPNPERIGSLVNSLMAFIDISSLHYMETANEGLOGRAFWCAGLS 647

```
Db 590 KIGSINWTRGYTIPSPERKSNPLNSLMGNFDIRSINQLETIKSSGPEPERELWLSGIA 649
QY 648 NEFHKSTKRBRGPHRLSGGYVIGGNLHTGSDKILSAFOQLGRDDYVVAANOQIVYG 707
Db 650 NEFYRDSMPTRHGRHRTSGSGLGATATTPAEDQLTFAPCOLFARDNNHJTGKNGDTYG 709
QY 708 GLLYYOHNETYISLPCKL-----RPCSLSYVPELIPVLFSGNLSTHTNDLTKTKYTY 761
Db 710 ASLYFHHTTEGLFDIANFLMWKATRAPWVLSEISQIIPLSDAKFSYIHTNHNKTYITDN 769
QY 762 PYTKGSGNDSPALDEGGRAPICLDESALFEQYMPMKLOFYVAHOEGEQTAEAREFG 821
Db 770 SIIRGSMRNDAPFADLGASLPFVIVPYLLKEVEPEFVKQYIYAHOQDFERYAEGRAFN 829
QY 822 SSRVLNLAIPGIRFDEKSDCODATVNLTLGYVVDLVRSPDCTTLTRISGDMKFEQTN 881
Db 830 KSELINVEIRIGTFEEDDSSEKGYDILLMTLIDATRRPKCOTSLIASDAMMAIGTN 889
QY 882 LARQALVLRAGNHFCNSNEAFSQSFELRGSSRNYNVDLGAKYOF 928
Db 890 LARQGSVRAANHFQVNPHEIFGQFAFEVRSRRNNTNLGSKCFE 936

RESULT 5
A81591
polymorphic membrane protein G family CP0307 [imported] - Chlamydia pneumoniae (stra
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: A81591
R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gelin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: A81591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <REA>
A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AAF38164.1; PID:g718923
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0307

Query Match 40.6%; Score 1940; DB 2; Length 930;
Best Local Similarity 43.7%; Pred. No. 1.3e-97;
Matches 416; Conservative 165; Mismatches 324; Indels 48; Gaps 17;

QY 1 MKSSFKPFVSTFAIFPLSM-IAT---ETVLDSSASFQGNKNGNFSVRESQEDAGTTTLF 56
Db 1 MKIPLHLKLISSLTVPILSLIATYGADASLPTDSFDGAGSFTPKSTADANGTNYVL 60
QY 57 KGNVTLENIPGTTATITKSCFNNTKGDLPFTGNGNSLLFQTVAGVAGAAVSSVVDKS 116
Db 61 SGNVYI-NDAGKGTALTGCCFTTETGDLFTTGKGYSFSEFNTYVAGSNAGAAA-STTADRA 118
QY 117 TTFIGFSSLSFIASPSSITTTGKAVSCSTGSLTKKNVSLSKNFSND--NGGAITA 173
Db 119 LFTTGFSNLSEIFAAPGTVAAGKSTLS-SAGALNLTGNGILFESQVNVSEANNNGAITA 177
QY 174 KTLSLGTMTSAFSENSTSSKKGAIQTSALITITGNOGEVSSDNTSSDSCGAITFEAS 233
Db 178 KTLISIGNNTSTFTSTNSAKRLGAGIYSSAAASISGNTQOLVPMNKGETGGGALFEAS 237
QY 234 VTISNAKVSFIDNKVTGAASSFTTGMGGAICAYKTSTDTKTLTGLNGOMLFFSNNTSTT 293
Db 238 SSTTQWSSLSLFSGNTATDAAG-----KGAITCEKTGETPPTLTISGNKSLTFAENSVYT 291
QY 294 AGGAIYVKLELASGLLTJFSRNSVNGTAPKGAITAIEDSGELSLADSGDIVPLGNTV 353
Db 292 OGGAICAHGHDLSAAGPTLFSNNRCNTAAGKGAITAIADSGSLSLSANQGDITFIGNTL 351
QY 354 TSTT-PGTRNSSIDLTGSAKMTALRSAGRAIFYDPI---TTGSSFTYTVDLKVNETPA 409
```

```
Db 352 TSTSAFTSRNMAIYLSGSAKITNLRAAGOSITFYDPIASNTTGA-----DVLITNQPOS 407
QY 410 DSALOYTGNIITFGEKLSSETAADSKNLTKLQVPTLSSGLSKHGVTLOQTQATQOA 469
Db 408 NSPLDYSGTIVSGEKLSDAEKAADNFTSILKOPALASGTALKGNVELDVNGFTQRE 467
QY 470 DSRLEMDVGTLEPAPTSIN--NLVINISIDGAKKAIENKATSKNLTLSGTITLDP 527
Db 468 GSTLLMQPGTKLK-ADTEAISLTKLVVDLSALEGNKSVEIETAGAKTKTTLTSPVFOOS 526
QY 528 TGTTFENHSLRNPOSYDILELKA-----GTVSTAIVPDPIYGEKPHYQGTG 578
Db 527 SCNFYFESHINQAFQPLVVFPAATAADYIDALLISVQGPPEP-----HYGQGHHE 580
QY 579 PIYWGCASTTATFNNTKGYIPNPERISGLVPSNLMANFIDISSIHYMETANEGLOGD 638
Db 581 ATWADITSAKSGTMTVWTGTGYNPERRASVVDPSLMASTDIRTLOQIMTSQANSIYQO 640
QY 639 RAFWAGLSLNEFHKSTKRRCGRPHRLSGGYVIGGNLHTGSDKILSAFOQLGRDDYFV 698
Db 641 RGLMASGTANFHHKDKSGTNOAFRHKSYGYVGGSADEPSENFVAFQOLGKDKDLFI 700
QY 699 AKNOGTVGGTLYYOHNETYISLPCKLRPC--SLSYVPELIPVLFSGNLSTHTNDLKT 756
Db 701 VENTSHNYLASLYLOHRAFLGILP--MPSFGSITDMLKDIPILINAOQLSYTKNDMDT 757
QY 757 KATTYTVAGSGNDSFALFEGGRAPICL-DSALFEQYMPMKLOFYVAHOEGEQT 815
Db 758 RTTSTYEDAGSWTNNSGALGLSLALYLPKAPFQGIYPLKQOAYSROONRESGA 817
QY 816 EAREFGSSRLVNLAPGIRFDEKSDCODATVNLTLGYVVDLVRSPDCTTLTRISGDSW 875
Db 818 EARAPEGDGLVNCISIPVGRLEKISDEKNNELISLAYIGDVYRRNPRRTSLMAGSAM 877
QY 876 KTFEYTLARQALVLRAGNHFCNSNEAFSQSFELRGSSRNYNVDLGAKYOF 928
Db 878 TSLCKNLARQALASAGSHLTLSPHYELSGEAAEYLRGSAHYNVDCGLRYSF 930

RESULT 6
D72078
polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: D72078
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D72078
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <ARN>
A:Cross-references: GB:AE001627; GB:AE001363; NID:g4376721; PIDN:AA018590.1; PID:g437
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_8

Query Match 40.6%; Score 1936; DB 2; Length 930;
Best Local Similarity 43.5%; Pred. No. 2.1e-97;
Matches 415; Conservative 165; Mismatches 325; Indels 48; Gaps 17;

QY 1 MKSSFKPFVSTFAIFPLSM-IAT---ETVLDSSASFQGNKNGNFSVRESQEDAGTTTLF 56
Db 1 MKIPLHLKLISSLTVPILSLIATYGADASLPTDSFDGAGSFTPKSTADANGTNYVL 60
QY 57 KGNVTLENIPGTTATITKSCFNNTKGDLPFTGNGNSLLFQTVADAGVAGAAVSSVVDKS 116
Db 61 SGNVYI-NDAGKGTALTGCCFTTETGDLFTTGKGYSFSEFNTYVAGSNAGAAA-STTAKRA 118
QY 117 TTFIGFSSLSFIASPSSITTTGKAVSCSTGSLTKKNVSLFSKNFSND--NGGAITA 173
```

[illegible]

A;Gene: pmp_11

Query Match	38.9%	Score 1855;	DB 2;	Length 928;
Best Local Similarity	42.7%;	Pred. No. 5.1e-93;		
Matches 405;	Conservative 171;	Mismatches 331;	Indels 42;	Gaps 20

QY	1	MXSFEKFEVFETFAIP--LSMITEFVLDSASFQDN-KNGNSVSEESQEDGATYLFK	57
Db	1	MKTSIPWLVSVYAFSCHLOSLANEELSLPDDSFNINIDSGTTTPKTS----	56
QY	58	GNVTLENIPTGTAIITKSCFNNTKGLDFTGNGNSLLFQTVDACTVAGAAVNSSVDKST	117
Db	57	GDVFEYFE-PGKGPTLSDSCFEKQTDNTDLTFGNGHSLTFEGDITAGHAGAAA-STTARKNL	114
QY	118	TEFIFSSLSFASJASGSSITTGKGAVSGCSTGSLSTKRVLSLFFKNSFETDNGATLTKLS	177
Db	115	TFSGFSLSDSSSTTIVTGGQLS--SAGVNLNENIRKLVAAGNESTAGGATKGSFL	173
QY	178	LGTGTMSEFSENTSSKKGAIQTSDDLITTGNOGEVSFSDNTSSDGAALFTEASVTIS	237
Db	174	LGTSGDALFENNSSSTKGAIAATTAGARIANTGYRFLSNIASTSGAIDDEGTSLIS	233
QY	238	NNAVSFELDKNVTAASSSTTGDMGGAICAKYSTDPKRVLLTGOMLFEFNNSTTACGA	297
Db	234	NNKFLYF----EENAKTT---GGAICNTKAGSGPELLISNNKTLIFASNAEYISGA	284
QY	298	IYVRKLELASGGLTFERNVNGCTARKGAAIAEDSEGSLSDSGDIFYLGNVTST-	356
Db	285	IHAKKLASSGGEFLERNVNSMT-PKGAISIDAGSELSTAEIGNITFVRRITLTTC	343
QY	357	-TPGTNSSSIDLGTSAKMTALRSAGAAIFYDPITTTGSSVTYTDVUKVETPADSLÖY	415
Db	344	STDTPKRAINIGSNGFETELRAKKNHTIFFEYDIT--SGTSSDVTKINNGSAGALNPY	401
QY	416	TGNITFEKGESEFEADSKMLTKLQPTLSGGLTSLKHGYLTQQAOTQOADSLEM	475
Db	402	QGTILFSGEITLADBLKAVADLKSFTLPQPSLSGSKLLÖKVTLBESTSQEASLIGM	461
QY	476	DVGTTL-PADTSTNNLVINISSIDGAKKAKIETKATSKNLTLSGTTITLLDPGTFEYEN	534
Db	462	DSGTTITTMAGSIITTMGINVDSLGKOPVSLTAKASKNKYVSGMLNIDIEGNYES	521
QY	535	HSLNPOSYDLLEKASGYTAS---PAVPPDPMGEKFNHYGOGTGPVWGASTT-	589
Db	522	HMFHDDQLFLSKTTVDADVDNTWDISSLIPVPAEDPNSEYEGOGQMN-VNMITTDATNT	580
QY	590	--ATFNMTKGYIPNPERIGSLVPSNLMAFIDISSLHYLMETANESLQGDRAFWCGLS	647
Db	581	KEATATYTKGFVSPERRKSALVCNTLMGFTDIRLOQVEIGATMEHKGFVWSSMT	640
QY	648	NEFHKDSKTRRGRHLSGGVATGCGNLHTCSDKLSAFCOLGGRDDYVAKNOGTVYG	707
Db	641	NELKLTGDENNKGRHHTSGGVITIGSHTPRKDDLTFTAFCHLEARDKDCITAHNSRTYG	700
QY	708	GTLVYQHNET----YLSL-PCKLRPASLYVPELIVLFESGNLSYTHTPNDLTKTKTYT	761
Db	701	GTLFFKHSHTLQPNYTLRLGAKPSESAIEKPREIPLADVOYFSFSHDNRRETHYSL	766
QY	762	PTVGSNGNDSFALFEGGRAPICL-DSALFEOYMPMKLQFYAHOEGKKEOGTEAREF	820
Db	761	PESGGSMSNECIAGIGLIDLPFVLSNHPLEKFTFIPOMKEMXYVSONSFESSDDRGF	820
QY	821	GSSRLVNLALPICIGIRPKESQCODA-FYNLTLAGTVLVSNDPCTTLRLISGSMKTFG	879
Db	821	SIGRLMLSTPVGAKF-VQGDIGDSYTYLDSGFEVDVYANNPQSTATLVMSDPDSMKIRG	879
QY	880	TNLARÖALVLRAGNHPCPNSFEAFQSFSFELNGSSNNYVWDJAKQOF	928
Db	880	GNLSRÖAFLLRGSSNNYVNSNCELFGEYAAELNGSSNNYVWDGKTJRF	928

RESULT 8

[illegible]

RESULT 10
C81591
Polymorphic membrane protein G family CP0309 [imported] - Chlamydothrix pneumoniae (strain
C:Species: Chlamydothrix pneumoniae, Chlamydothrix pneumoniae
C:Date: 31-Mar-2000 #sequence-revision 31-Mar-2000 #text-change 11-May-2000
C:Accession: C81591
R:Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.C.; Dodson, R.; Gunn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
M.; et al. Genome sequences of Chlamydothrix trachomatis Mopn and Chlamydothrix pneumoniae AR39.
A:Title: Genome sequences of Chlamydothrix trachomatis Mopn and Chlamydothrix pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: C81591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1276 <REA>
A:CROSS-References: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AA838166.1; PID:g7189222
A:Experimental source: Strain AK39, HL cells
C:Genetics:
C:Gene: CP0309

Query March 30.2%; Score 1444; DB 2; Length 1276;
Best Local Similarity 29.5%; Pred. No. 1.5e-70;
Matches 382; Conservative 157; Mismatches 370; Indels 386; Gaps 26;
QY 1 MKSPPKVFVFSTFAIFPL-SMATETVYLDSSASFDGNKNGN--ESVRSOEDAGTYLFK 57
DB 1 MKYSPMLITLSALVFLHPMLAANTDLSSSDNYENGSSGSAAFKAPKPTSDASGTYYLLT 60
QY 58 GNVLTLENIPGTATIT---KSCFNNTKGDITFTTGNSLLTFOTVDAGIVAGAAVSSVVD 114
DB 61 SDVSTTAV---SALTTPADKSCFTNTGALSFVGDHDSIVLOTI-ALTHDGAALNNT--N 113

Oy	115	KSTFFIFGSSLSFIASGSSITTCOKAVSST-----GSLSTRKNVSLSEKRFSPNDNGA	170
Db	114	TALFSGSSLLIDSAFATGTSGGKAI-CVTEEGGTATFDMAVLTLOKNTSEKDGA	172
Oy	171	ITAKTLSTGTTMALSEUNTSSKKGAIDTSDALITGNOGEVSFSDTSSDSGAIF-	229
Db	173	VSASIDAKTTTAALDONTSTRNGCALCSTANTTVQGNISGVTYSNMTAIDKGGIYS	232
Oy	230	-----TEASVISNNAKVSFIDNKVTG-ASSSTTGD	260
Db	233	KEKDSTLDANTGVTFEKSMTAKTGAWSSDDNLATGNTOVLEQEKKTGSAOANNPEG	292
Oy	261	SGGALCAV-KTSD-----TKVLTGNOMLLSNNTS	291
Db	293	CGGALCCYATADKTCGLAISONEMSFTSNNTTANGAIIAYATKCTLDNTTLTFEDONTA	352
Oy	292	T-----	292
Db	353	TAGCGAIYETEDFSLKSGTGVTFSINTAKTGALYSKGNSSLTGNTLNLFSGNKATG	412
Oy	293	-----	292
Db	413	PSNSSANOEGCGAIIISLESASVSTKKGLIEDNENVSLSGNTATVSGCAIATKCALH	472
Oy	293	-----	292
Db	473	GNTTLTLDGNAETAGAIYETEDFTLGISGTIVFSTINTAKTAGALHTKONTSTFTANK	532
Oy	293	-----	292
Db	533	ALVSGNSATATATTDTDOEGCGAIIICNISEDIATKSLTTLTENESLFINNTAKRSQG	592
Oy	293	-----TAGAIYVKKLELASGGLTFBSNVSNGSTAKGAI	329
Db	593	GIYAPKCVIISGESINFDGNTAETSGGALSKNLSTIANGAPVSEFTNNS--GG--KGGAI	647
Oy	330	AIEDSGELSLASDGDIVFLGNTVSTPTGPNSSSIDLGTSAKMTALRSAGRAIIFYDP	389
Db	648	YIADSGELSLAIDGIDITFGSNATGEGTSPN-SIHLGAKAKITLAAAPHITLYFDP	705
Oy	330	IT-----TGSSITTVD-----VLKAVETPADSLQYTGNITTEGKL	426
Db	706	ITMEPAPASGGITEELVINPVYKAIYPPPOKPNPISASVPVAPVAPNPMPTGIVISSGKL	765
Oy	427	SETEPAAOSKNTSKLDOPVTLSCGTELSLKHGVLTQAFTOOADSRLXNDVGTLEPADT	486
Db	766	PSQASIPANTTILNKINLAGGNVYLNKGAILOYSTTQOPDSIVFEMDADTILETTT	825
Oy	487	ST-----INNLVINISSIDGAKKAKIETKASKNLTLSGTLTLLDPGTGFEYENHSLRNP	540
Db	826	NNTGCSIDLKNTLSYNIDALDALKRMITIYANVSTISGELKISGDLKFHNNEGSEFYANDGELKAN	885
Oy	541	QSYDILELKA-SGTVSTAVTPDPPINGEKFXHVGOGTGW--PIVNGTGASTATFENMKT	597
Db	886	LNLFLLDLSSTSGVINDPNIIPSSMAAPDYGQSSMTLVYKV-GAGGKATVLVAEMQAL	944
Oy	598	GIYENPERIGSLVPSNLMAAFIDISSLHYLMETANBGLDGRAFWCAGLSNFFRHKDSTK	657
Db	945	GYTFKPELBAITLVENSLMAYVNIHSIOEIATAMSDAPSHPOIGWIGIGNAFHQDKQE	1004
Oy	658	RGRFRHLISGGVYIGGNLHTOSDKILSAAFCQLFGRBODYFAKNOGTYGGTLYVNOHET	717
Db	1005	NAGFRLLISRGIVGSSMTTPOEYTFPAVAASQLEKGSKRDIVYDILSOYVAGSLCAO-SSY	1063
Oy	718	YISLPCKLRPCSLSV---PREIPLYEGLSNLSYTHTDMLTKYTYTPVYGSNGNSF	773
Db	1064	VIFPHSLSRHRYLVSKVPELPGETPLVLYNGOVSYGGNHNMMTTKLANTNOGKSDMDNSF	1122
Oy	774	ALFEGGARPICLDESALEDOYMPFKLODYVAHQSEFKOGSTEARBFGSSRLVNLADPIG	833
Db	1124	AVEGGSSLPVDLNNRYL-TSYSPYVLQOVYVAKQOFQVADAPRIFDASHLUNVANSIPMG	1182


```

0Y 336 TPGNRSSIDGTAATLSAAGRAIYFEDPTTSSSTVTVLVNENPASALOY 415
0Y 351 SLFPLN-GIHLLOAKKLKLOARNGYSIEFDPTT--SEADGSTQOLINGDPKK--EY 404
0Y 416 TGNIFTEGKLSETEADSKNLTSLKLOPVLTVLGGTSLKHGVTLOTOAFTQOASRLM 475
Db 405 TGTLFSGEK--SLANDPROFKSTIPQVNVLSAGYLVIKBEAEVTEKFPQPSHLVL 461
0Y 416 DVGITL-EPADTSTNNLVINISIDGAKKAK-IETKATSKNLTSGRTTLIDPTCTYE 533
Db 462 DLGTLILASKEDIATGTLAIDIDSSSSSTAIVIAMTANKOISTVDEIELISPGNAYE 521
0Y 534 NNSLKNPOSYDLEIK--ASGTVTSTA-----VPRDPIMGKEFHGXGTWPIYW-GTG 585
Db 522 DLRMNSOTFFPLSLTEPGAGSVTVTAGDPLVSP-----HYFGGNM-KLMTGTG 572
0Y 586 ASTATFNVTKTGYIPNPERIGSLVPNSLIMNAFIDISSLHYLMETANGLAGRAFWCAG 645
Db 573 -NKVEEFFMDKINYPRPREKCNLVPNILMGNAVDRSLMQOETHAESLQJDRBLMIDG 631
0Y 646 LSNFHKASTKTRRFRHLISGGYVIGNLHTCSDKILSAFCQLFGRPDRFVANKOCTV 705
Db 632 IGNEFHVASSEDNIRYRHRNSGGYVLSVNNELIPKHYTSMASFOLFSRDKDVAVSNNEYRM 691
0Y 706 YGGLTYQHNER-----YISLPCKLRPCSLSYVTEIP-VLFSNLSYTHDNDLTKRY 758
Db 692 YLGSYLYQYTTSLGNIERFASRNPVNVNGLISRLFLQNLMIHFELCAYGAHNTDMKXDY 751
0Y 759 TTYPVKSMGDSFALFEGGRAPICLDESA-LFEQYMPFKLQVYVAHOEGFKDQTEA 817
Db 752 ANFPVKKSMWRNCNAIECGGSMPLVFEENGRLFGALIPFKMLQVLVYAKGDFKTTIDG 811
0Y 818 REFSSRLVNTALPIGIRFDKESDCODATYNNLTIGYTVDLVRSNDPCTTIRISDSMKT 877
Db 812 RRFSSGSLTISIVPLGIRFEKTLAKQDVLVYDFRSFYIDPIFRKDPSCBAALVSDSMVLV 871
0Y 878 FGTNLARQALVLRAGNHCFNENFEAFSOFSELMGSSRNXYVVDGAIYOF 928
Db 872 PAAHVSRAHFAVGSGTGRYHFENDYTELLCRGSLTECPHARRNININGSKFRF 922

```

```

RESULT      13
B72078
polymorphic outer membrane protein g/i family - Chlamydophiila pneumoniae (strain CWL029)
C:Species: Chlamydophiila pneumoniae, Chlamydophiila pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: B72078
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Native Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydophiila pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: B72078
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1*1407 <ARN>
A:Cross-references: GB:AE001627; GE:AE001363; NID:q4376721; PIDN:ADI8588.1; PID:q4376721
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_6

Query Match      29.0%; Score 1383.5; DB 2; Length 1407;
Best Local Similarity 27.3%; Pred. No. 3.3e-67;
Matches 388; Conservative 156; Mismatches 370; Indels 507; Gaps 28;

Oy      1 MKSSPKFVFESFAFPL-SMIATEVLDDSSAFPGNKNGN--FSVRSDOEDATTYLEK 57
        1 MKYSLPMLITSSALVFSLHPILMAANTDLSSDNTEGSSGSAFTAKETSDASGTTITLT 60
Db
Oy      58 GNVLTENIPGTGTAIT---KSCFENNTKGDLTFTFGNGNSILTFQVDAGTVAGAANVSSVVD 114
        61 SVSTSTTV-----SATTPADKSCFTNTGALSFVGADHSIVLOTI-ALVHDSGAINTT--N 113

```

Qy	115	KSTFIFGSSLSIFASGSSITTGKAVCSST-----GSLTRKNVSLFPRKPFSDNGA	170
Dd	114	TALSFSGFSLILDSAPATGSGKGAI-CVYTBEGITATFDMAVYLQAKTSEKDGA	172
Qy	171	ITAKTLSTLGTMSALFSENTSSKKGAIOTSDALITGNOGEVFSFSDTSSDGAIF-	229
Dd	173	VSAYSIDAKTTTALLDONTSTNGCALCSTANTTVQSGSVTFSSNTANDKCGIYS	232
Qy	230	-----PEASTISNNKAVSTIDNKVNG-ASSSTTODM	260
Dd	233	KEKOSTLDANTGVTFEKSNTAKTGGANSSDDNLALGTNTQVLEOENKTTGSAAOANBPG	292
Qy	261	SGAICAY-KTSTD-----	273
Dd	293	CGGAICCYLATATDKTGLTAISONEMSFSTNTTANGAIYATKCTLDNTLTLPDONTA	352
Qy	274	-----	273
Dd	333	TAGCGGAIYETEDFSLKSGSTGVTFSTNTAKTGALYKSGNSLTGNTLLFSGNKATG	412
Qy	274	-----TKVELT	279
Dd	413	PSNNSANGEGCGAILAFIDSGVSOKTGSLIANNQEVLSJTSMAATVSGAIYATKCTLT	472
Qy	280	GNOMILFSNNTSTTAGAIYVK-----	302
Dd	473	GNGSLTEPDGNTAGSGAIYETEDFTLGTGTGVTFSTNTAKTGALYSKGNNSLSGNT	532
Qy	303	-----LELAS	307
Dd	533	NLFSGNKATGPSNNSANQCGGAILFLELSASVSTKKGMLTMDENENLSGNTATYSG	592
Qy	308	-----GCLTL-----	312
Dd	593	GAIYATKCALHGNTTLLFPDNTAETAGAIYETEDFTLGTGTGVTFSTNTAKTAGALH	652
Qy	313	-----FSRNSV-----	318
Dd	653	TKGNTSFFKNALVYFSGNSKATATATTTTDOEGCGAILCNISBDAIATSLTLENESIS	712
Qy	319	-----NCGTAP-----	324
Dd	713	FINNTAKRSGGAIYAPKCVISGSEBSINFDPNTAETSGAIYSKNLSITANGVSEFTNSG	772
Qy	325	-KGGAIEDSGELSLSDSGDIVFLGNTVSTPTGTSNSIDLGTSAKMTLRSAAGRA	383
Dd	773	GKGAIYIADSGELSLAIDGDIFFSGNATBGTSPN-SHLGGAITITLAAAPGHT	830
Qy	384	IYFVDPIT--TGSSTVTVD-----VLKIVETPADSALOYTGNII	420
Dd	831	IYFDPITMEAPAGSGIIEELIYINPVKAIYVPPQPKNGPIASVPVPAVAPANPMTGTIV	890
Qy	421	FTGEKLSLEAADSKNLSKLDLPVYLSGTSLSKGVYLLQOAFQOADSLENDVCTT	480
Dd	891	FSSGGLPQSDASIPANTTITLQKINLAGGNVLLKGAILOVSEFQOQDSVFEVDAGCT	950
Qy	481	LEPADTST-----INMLVINISSIDGAKKAKIETAKTSKNLISGITIILDPGTGFYEN	534
Dd	951	LETTTNTNTOGSDIDKMLSVNLDALDQKRNITIAVNSTSGGLISDGLFHNNEGSEFDN	1010
Qy	535	HSLRNPOSTYDILEKA-SGVYTSATVPDPIMGEKHYGOGTWC--PIYMGTCASTAT	591
Dd	1011	PGLNANLNPFLDLSSTSGVNUINDDFNPFPSSMAADYGOGSMTLLVPKV--GAGKRVLY	1065
Qy	592	FNNRTGTIPNPERIGSLVJNSLMAAFIDISSLHYMETANGLOGDRFVFCAGLSNFFH	651
Dd	1070	AEOMALGTTPKPELRATLYVNSLMNAYVNIHSIQOIAIATMDAPSHPOIWTGIGGNFH	1122
Qy	652	KDSTKTRRGGRPHLSGVIYIGNLHTCSDKILSAFCOLFGRDRDYFVAKNOGTWYGCFLY	711
Dd	1130	QDKQRENAWGFILSRGIYGVGSMTTPOEYVFAVAFSQLFKSKRDYVVSIDIKQYVAGSILC	1189

[illegible]

QY 712 YOH:::YVLSLPCXKLRPCSLSY----PTEIVFVFSGMLSYHTNDKATKTYTTPYVKG 767

Db 1190 AQ-SSYVPLPLHSSLRHRLSVKLPBEPTEPLVHGVSTGRNHHNNTTTLANNTOCKSD 1248

QY 768 WGNDSFALFEFGRAPICLDESALFEQVYKPEFKLOFYVAHOGEFKEQCTEAREFCSSRLVN 827

Db 1249 WDSHFANAEVGSGLPVDLNTYRL--TSYSPYKLOVYSVNOKGPEVAAADRIEDASHLVN 1307

QY 828 LALPIGIFEDKESDQDQATYNLTGTYVDLVRSNPDCTTTLRISGDSWKTFGTNLAROL 887

Db 1308 VSIPIGLTFEKHESAKPSPALLLTIGVANDARDHPHCLTSL-TNGTSMSTFATWLSHOAF 1366

QY 888 VLKAGNHFCFNSNFEAFSQSFELRGSSRNVDLGAKYOF 928

Db 1367 FAEASGHKLILHGLDCEFASSGCELRSSSRSSRNANCGRIYTSF 1407

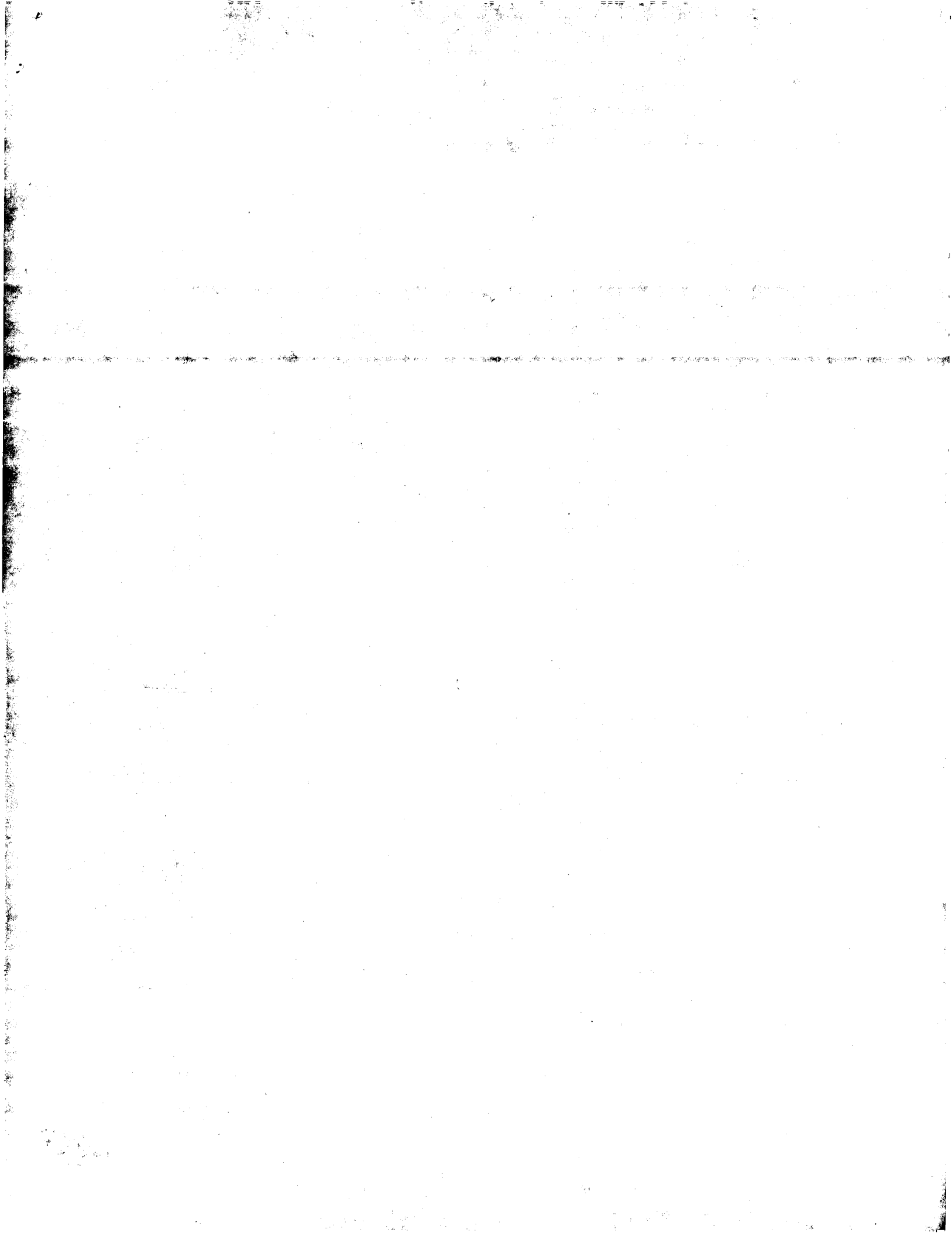
RESULT 14
F72076
polymorphic outer membrane protein g family - Chlamydophila pneumoniae (strain CWL029)
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: F72076
C:Kaltman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: F72076
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-973 <ARN>
A:Cross-references: GB:AE001629, GB:AE001363; NID:q4376734; PIDN:AAD18595.1; PID:q4376734
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_13

Query Match	28.9%	Score 1377.5;	DB 2;	Length 973;
Best Local Similarity	34.8%;	Pred. No. 4.1e-67;		
Matches 350;	Conservative 153;	Mismatches 391;	Indels 113;	Gaps 25;

[illegible][illegible]

RESULT 15
C81593
Polymorphic membrane protein G family CP0299 [imported] - Chlamydothilla pneumoniae (s
C:Species: Chlamydothilla pneumoniae, Chlamyda pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: C81593
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldenberg, J.F.; White, O.; Hickey
C.: Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: A81500; M01D:20150255
A:Accession: C81593
A:Status: preliminary
A:Molecule_type: DNA
A:Residues: 1-995 <RE>
A:Cross-References: GB:AE002191; GB:AE002161; NID:g7189216; PIDN:AAF38156.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0299

	Query Match	28.9%	Score 1377.5;	DB 2:	Length 995;
	Best Local Similarity	34.8%;	Pred. No. 4.2e-67;		
	Matches 350;	Conservative 153;	Mismatches 391;	Indels 113;	Gaps 25;
OY	1 MKSSPKFVFST-----PAIFPLMIATETVLDSSASFQMKNGNFVSRESDGADGYTL 55 : : :				
Dd	23 MKTSTRKRLISTTLAPCA-----SIAFLVEIIMPSBENDSGSKLFPPTTTSDPGITLCI 78 : : : :				
OY	56 FKGAVLTLENIGTGTAIRKSCFNNTKGDILFTGNNSLLFOIVDAGTVAGAAYNVSVYDK 115 : : : :				
Dd	79 FSGDLIANLDMALSRSTSFCFSNAGALQLLGKGVGFSLNI--RSSADGAAISSVIINQN 137 : : : :				
OY	116 -----STTIIGRSSLF-----IASPSSITTTGKGAVSCSTGSLTKNVSLFKNFS 164 : : : : :				
Dd	138 PELCPLEPSGSOMFDNCESLIDTSPTSASANIIPHASAIYAATTSPMFI-FNNNSILFEQYRS 196 : : : : :				




```
Db 240 C-AGGAIFSPICSLTGRGNIVFYNNRCKKNVETASSEAS---DGCAL-----KVTTRLD 290
Oy 278 LTGNQ-MLFFSNNTSTTACGAIYVKKLELASGGLTFPSNSVNGTAPGAGAIADDSGE 336
Db 291 VGNRGRIRFESDNIRKNYGAIAYPVLTVDNGPTFENNINNANV---KGGAIYIDGTSEN 346
Oy 337 LSLSDSGDIVELGNTVSTT--PGTN-----RSSIDLGSAKMTALRSAGRAIYF 387
Db 347 SKTISADRAHIIINENIYVITNANGTSTANPPRRNATIVASSGELILGAGSSQMLIYF 406
Oy 388 DPTTGSSTTVTVDLKVNETPADSALOYTGNIIFTEGKLESTEDADSKMLTSKLOPVL 447
Db 407 DEIVSNAGVSYSENK-----EADQ-----TGSVFCATVNSAD-FHQRMLOQKTPAPLTL 457
Oy 448 SGGTSLKRGVLTQATQOAD-----SLEMGVGTLEPADSTINNLYINIS 497
Db 458 SNGFLCIEDHAQLYNRFOTGGVYSLGNGAVLSQYKNGTGDSSASNA-SITLKHIGLINS 516
Oy 498 SI--DGAK--KAKIETKATSKNLT--LSGTITLDPDTGFYENHSLRNPQSDILELKAS 551
Db 517 SILKSGAEIPLMVEPTNSNNYTDATATFSLDVKLSLIDDYGSPESTDLTHALLS 576
Oy 552 GTVTSTAVPD-PINGEKF-----HYGQGTWGPVWG-----TGASTATAT-- 591
Db 577 QPMLSTSEASDNLQLOSENIDFSGLVNPHYGMOGLW---TWGNAKTQDPPASSATITDPQ 633
Oy 592 -----FWMTKGYIPNPERIGSLVPSLW-NAFIDISLHMLMTANELOGDNA 640
Db 634 KNRHRHRLTLTLWPAGYVPSPKHRSPLIANTLWMLATSLKNSALITPSG---HP 689
Oy 641 FW---CAGLSNFFHKDSTTRRGFRHLSGQY---VIGNLATCSDKILSAFCOLFGDR 694
Db 690 FNGITGGGLGMVYDDPRNRHHPGFHRSSGYSAGMTAGTHTFSLK-----TSQITTKLN 744
Oy 695 DYFVAKNOSTYVGGILYOHNETYISLPCKLRPCSIYPTETIPLVFS---GNLSTYTHD 751
Db 745 ERYV-AKNV---VSSKNYSCQGEMLFSL-----OEGFLITLVGLYSYGDHNCHEFYTQ 793
Oy 752 NDLKRYTTY--PTYKSGNDSPALFEEGGRAPICIDESALREQVYMPFKLOFYVAHOG 809
Db 794 GENLTSOGTFRSQTWGGAFFDLPMKPFSGTHILT-----APFLGALGITSLSH 843
Oy 810 FKEQGTAREFGS-SRLVNLALPIGIRFEDKSCODAT-----YNLLGTATVDLVRSPD 863
Db 844 FTEVGAYPRSFSTKTPPLINVLVPIGVK---GSEFMATRQPAWYELAYQVYLRQEBG 899
Oy 864 CTTTIRISGDSKRTGTNLARQALVLRAGN-----HFCFNSNFEAFSOFSELRG 913
Db 900 IAAOLASKGIMFGSGSPSSRHAMSXYKISQOTQPLSWTLHFQYH-GFYSSSTFCNYLNG 958

RESULT 2
OMPF_CHLTR STANDARD: PRT: 1034 AA.
AC P38008; 084878;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN P PRECURSOR.
GN PMP OR CT870.
OC Chlamydia trachomatis.
OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RC STRAIN=D/W-3/CX.
RX MEDLINE; 99000809.
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.:
RT "Genome sequence of an obligate intracellular pathogen of humans:
RL Chlamydia trachomatis."
Science 282:754-759(1998).
```

```
RM [2]
RC SEQUENCE OF 26-35.
RP STRAIN=L2/434/BU.
RA Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA Comanducci M., Christlanen G., Birkelund S., Viret E., Ratti G.,
RA Pallini V.:
RL Submitted (SEP-1994) to the SWISS-PROT data bank.
CC -i- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES).
CC
CC This SWISS PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE001360; AAC68468.1; -.
KW Outer membrane; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1034 PUTATIVE OUTER MEMBRANE PROTEIN F.
FT CONFLICT 34 35 RR->FH (IN REF. 2).
SQ SEQUENCE 1034 AA; 112392 MW; 445FF4C3D463AE7 CRC64;
```

Query Match 10.3%; Score 490; DB 1; Length 1034;
Best local similarity 23.5%; Pred. No. 6, 1e-20;
Matches 258; Conservative 143; Mismatches 428; Indels 268; Gaps 49;

```
Oy 5 PPKVEFS---TEAFPLSMATETVLDSSASFDGKNKGNFSRESQDAGTYLTKGNV 60
Db 33 FRFFFSREIOFVDPDPAFLTAQNVLS--NLQNGTGACTIONTQ---TQIFSNV 86
Oy 61 TLENPGGTATKSCFNNTKGDLLFTGNGNSLFEQYDAGTVAGAAVSSVVDKSTFEI 120
Db 87 NT-----TADSGAFPMVTTSTFASDNANLFCNNYCTHNKGGALRSG--GPIFL 136
Oy 121 GFSLSF---IASPSSITTGK-----GAVSGTGLSTLTKVNSLFSKNFTDNGA 170
Db 137 NNQDPLFPNNISAGKAYGCTGDHNEKNKGALYAT--ITLGNRTLAFINMSDCCGA 194
Oy 171 ITAKT-LSLTGTMALFSEN-----TSSKKGAI-QTSDALTITGNGEVSFS 217
Db 195 ISADQIISITTVKILFENNHTLNIPIYTOAENNAKRCISRDLCISNNSGPIVF 254
Oy 218 DNTSDSGAAFTESAVTISNNAKVSFIDNKVTGASSSTGDMSGAICAKYTSRDTVT 277
Db 255 YN-QCGKGAISATRCVIDNKKERTIFSNSSLGWSOSSAS--NGAL-----OTTGFT 307
Oy 278 LTGNQ-MLFFSNNTSTTACGAIYVKKLELASGGLTFPSNSVNGTAT-----PKGAIAT 331
Db 308 LRNKGSIYFDNSTATHAGAINCYIDIRDNGPYFLNNSAAMGAARNLKPR----- 361
Oy 332 EDGSELSADSGDIVELGNTVSTTBT-----NNSIDLGSAKMTALRSAG 381
Db 362 --SATNYIHTGTGDIV-NNNVFETLDGNLLGKRKLFIINNETPYT---LSIGAKKD 414
Oy 382 RAIYFYD-----PITGSSSTVTVDLKVN-ETPADSALOYTGNIIFTEGKL 426
Db 415 TRIIFYDLFQWERKKNENSNPPSPSTSRNITT---VAPET-----EFGAVVFSTYQM 464
Oy 427 SETEADSKNLTSK-----LLQPYTLSGTSLKRGVTLQ--TQATQOADSREMDVGT 479
Db 465 S-----SDIRTLMGKEHNRIKEAPTLTKGTIALIDDALELFNPFQNPSTSLALSGA 520
Oy 480 TLEPADTSTIN---NLVINISSIDGAKKAKIE-----TATSKNLTLSGTITLLDPTG 529
Db 521 TLTVGKHGKLNITNLGVLPIILKEGKSPCIRVNPQDMQNTGQTPSSSTSSISPMI 580
Oy 530 TFEYENHSL--RNPOS-VQILEL---KAGIVTSTAVVPPDIMGKF-----HY 571
Db 581 IFNGRLSTVDENYESVDSDMLSRKAQQLILSTETITNDGOLDNSNMOSSLNTSLSPHY 640
```

QY 572 GYGTGMPW-----582
 DB 641 GYGLMTP-WMTTYYTINNNSAPTSAIEOKTSEFTPSMTTASIPNIKASA 699
 QY 583 -GTGAST-----TATFNMTKTYIPNPERIGSLVNSL---WNAFIDISSIHY 626
 DB 700 GSGSGSASNSGEVITTKHTLVNNAIPYGIYDPIRBDGLINSLVHSGRNMTGLRSL-- 757
 QY 627 LMETANGLGDRAPWCAGLSNFFHKDSTKTRRGFRHLSSGYVI---GGLNHTGSDKI 681
 DB 758 LPDMSMELGQ-----AATLFTKOOKRLSYHSGYSASKGYTSSOASGAHGH----- 805
 QY 682 LSAFQCLFGRDRPYFAKN-----QGYVGGTLYOHNEHYIILPCKLRCSLSYPT 736
 DB 806 ---KFLSFSSOSSDKMEKETNNRLSSRYIYLSALCFE-----PMF 843
 QY 737 IYVLFSGNLSYHTDNDLTKYTYTPYVKGWGNDSFALFEGRAPIC-LDESALFEOYM 795
 DB 844 DRILIGAACNVCYTHMNRSPYGTAKSKGFHSTTL---GASLQELDMSPLMSIM 898
 QY 796 --PPMKIQFYVAHOEGKEGCTEAREFGSSRL-VNLALPIGRDKESDC-QDATYMLTL 851
 DB 899 LTPPAQLFRTPEPASIRESDIARFLTEQAHTRAVSPGIGKAYSDDTWPILSWEML 958
 QY 852 GYTVDLRSPDCTTTRISGDSWKTFTGNLAROALVLRAGNFCFNSNEAFSGSFEL 911
 DB 959 AYOPTLWKRPRLNTLILONNGSWTNTPLAKHSFYGR-GSHSLKFSHLKLFANYOAEV 1017
 QY 912 RGSRRNVLDGAKYOF 928
 DB 1018 ATSTVSHYNAGALVF 1034

RESULT 3
 190K_RICRI STANDARD; PRT: 2249 AA.
 AC P15921;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-R.
 RX MEDLINE: 90354033.
 RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.J.,
 RT "A protective protein antigen of Rickettsia rickettsii has tandemly
 RT repeated, near-identical sequences.";
 RL Infect. Immun. 58:2760-2769(1990).
 CC -FUNCTION: ELICITS PROTECTIVE IMMUNITY.
 CC -PTM: GLYCOSYLATED (POSSIBLE).
 CC -DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF
 CC AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.
 CC -This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M31227; AAA26380.1; -.
 DR PIR: A41477; A41477.
 KW Antigen; Repeat; Signal; S-layer; Glycoprotein.
 FT SIGNAL 1 28
 FT CHAIN 29 2249 POTENTIAL.
 FT DOMAIN 212 1180 190 KDA ANTIGEN.
 FT REPEAT 212 286 13 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 287 358 A (TYPE I).
 FT B (TYPE II).

FT REPEAT 359 430 C (TYPE II).
 FT REPEAT 431 505 D (TYPE I).
 FT REPEAT 506 577 E (TYPE II).
 FT REPEAT 578 652 F (TYPE I).
 FT REPEAT 653 724 G (TYPE II).
 FT REPEAT 725 799 H (TYPE I).
 FT REPEAT 800 874 I (TYPE I).
 FT REPEAT 875 949 J (TYPE I).
 FT REPEAT 950 1021 K (TYPE II).
 FT REPEAT 1022 1093 L (TYPE II).
 FT REPEAT 1094 1165 M (TYPE II).
 FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
 SQ SEQUENCE 2249 AA; 224333 MM; A9D6646C089DF087 CRC64.
 Query Match 5.6%; Score 269; DB 1; Length 2249;
 Best Local Similarity 22.9%; Pred. No. 3.4e-07;
 Matches 203; Conservative 100; Mismatches 324; Indels 258; Gaps 43;
 QY 60 VTLENIGCTAITKSCFNNTKGDITFTGN-----GNSLFFQYDAGTVAGAAVNSSV 112
 DB 396 VTEFTN-PVVVTGAIDNTGNANNNGIVFTGNSVTYGDIGNALATVNVG--AGTATLGA 452
 QY 113 VDKSTT-----FIGSSLSFIASPGSSITG-----KGAVSCSGSLSLTKNV 155
 DB 453 VIKATTKLTNAASVLTLTNANAVLTGALDNTTGGDNVGLNGLALSOYTGINGNTSL 512
 QY 156 SLF-FSKNFTDNGALITAKTLTLTGTTMSALFS-----ENTSCKKGAIGTSDAL 205
 DB 513 ATISVAGATATLGAVIKATTKLTDAASAVKFTNPVVVTGALDNTGNANNNGIVFTGNS 572
 QY 206 TITGNOGE-----VSSDNTSSDGAAT-----FTEA-----SVTISNAKSFIDN-- 247
 DB 573 TVTGDIGNTSLATISVAGATATLGAVIKATTKLTNAASVLTLTNANAVLTGALDNTT 632
 QY 248 -----KVTGASSSTTGDM-----GGAICAYKSTDTKVTILGNQ 282
 DB 633 GGDNVGLNGLALSOYTGIDIGNTSLATISVAGATATLGAVI---KATTKKIT-NAVS 688
 QY 283 MLFSNNTSTTACAIYKLELASGGLTFSRNS-----VNGTAPKG 326
 DB 689 AVKFTNPVVVT--GAI--DSTGNANNNGIVFTGNSVTYGDIGNTALVNVGATATLG 744
 QY 327 GAIAIEDSGELS-----LSADSGDIFELGNTVSTTPGTRRSI-----DLGTS 370
 DB 745 GAVIKATTKLTNAASVLTLTNANAVLTG- AIDNTTGGDNVGLNGLALSOYTGIDIGNT 803
 QY 371 AKMTALRSAGRAIYFYDPIITGSSSTTVTDVLKVNETPADSALOYT-GNIIFTGKLESET 429
 DB 804 NSLATISVAGTA-----TLGAVIKATTKLT--TNAASVLTLTNANAVLTG-AVDNT 853
 QY 430 EAADSKRLTSLKLPVTLSCGSLTKHGVLTQATQOADSLEMDVGTLEPATDSTI 489
 DB 854 TGGDNV-----GVNLN-NGALSOYTGIDIGNTSLATISVG-----AGTATL 893
 QY 490 NNIVINISSIDGAKKAKIETKATSKNLTLSGTI---TLDPTGTFEYENHSLR-----N 539
 DB 894 GGAVIKATTKLTNAASVLT-LTNANAVLTGALDNTTGGDNVGLNGLALSOYTGIDIGN 952
 QY 540 POSYDILELKASGTVT-----STAVTPDPIMGEKPHYGYGTGWPIWGTGA----- 586
 DB 953 TNSLATISVGA-GTATLGAVIKATTKLTLDASAVKF-----TNPVV-VTGALDNTG 1003
 QY 587 -----STTAFENMTKTYIPNPERI-----GSLVNSLMAAFIDISSIHYLM 628
 DB 1004 NANNNGIVFTGNSVTGNVGNLTALVNVGAGLLOVGGVAAANINLTLDNASAVFTFN 1063
 QY 629 ETANEGLOGDRAPWCAGLSN---FFHKDSTKTRRGFRHLSSGYVIGNLHTGSDKILSA 684
 DB 1064 PVVVTG---AIDNTGNANNNGIVFTGNSVT-----GNV--GNTNALATVNVGA 1107
 QY 685 AFQCLGRDRPYFAKNQGYVGGTLYOHNEHYIILPCKLRCSLSYPTLPPVLFSGN 744

DB 1108 GLOVOG-----GVKANTINLTJDN-----ASAVFTNPVVVTGA 1142

QY 745 LSVTHDNDLKTYYTPYVKGSGMDS-----FALEGG 779

DB 1143 IDNTGNANGIVTFTGNSVTYTGIDGNTMALATVYVAGAGITLLOAG 1187

RESULT 4

SLAP_CAUCR STANDARD; PRT: 1025 AA.

AC P35828; Q046015; (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 30, Last sequence update)

DT 01-OCT-1994 (Rel. 30, Last annotation update)

DE 30-MAY-2000 (Rel. 39, Last annotation update)

GN S-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).

OS R5AA.

OC Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21: 646-660 AND 1021-1025.

RC STRAIN=ATCC 19089 / CB15;

RX MEDLINE: 93007489.

RA Gilchrist A., Fisher J.A., Smit J.K.;

RT "Nucleotide sequence analysis of the gene encoding the Caulobacter crescentus paracrystalline surface layer protein.";

RL Can. J. Microbiol. 38:193-202(1992).

RN [2]

RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.

RC STRAIN=ATCC 19089 / CB15;

RX MEDLINE: 89008089.

RA Fisher J.A., Smit J.K., Agabian N.;

RT "Transcriptional analysis of the major surface array gene of Caulobacter crescentus.";

RL J. Bacteriol. 170:4706-4713(1988).

RN [3]

RP CHARACTERIZATION.

RC STRAIN=CB15;

RX MEDLINE: 98292737.

RA Awram P., Smit J.K.;

RT "The Caulobacter crescentus paracrystalline S-layer protein is secreted by an ABC transporter (type I) secretion apparatus.";

RL J. Bacteriol. 180:3062-3069(1998).

CC -I- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.

CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.

CC -I- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AF062345; AAC38665.1; -

DR HSSP: P01549; 2MCW.

DR INTERPRO: IPR001343; -

DR PFAM: PF00353; hemolysinCabinid; 1.

DR PRINTS: PR00313; CABNDNGRPT.

KW Cell wall; S-layer; Calcium-binding.

FT INIT_MET 0

FT SEQUENCE 1025 AA; 98209 MW; AFC8B519820B1A5F CRC64;

Query Match 5.1%; Score 244.5; DB 1; Length 1025;

Best Local Similarity 24.2%; Pred. No. 2,6e-06;

Matches 179; Conservative 86; Mismatches 257; Indels 217; Gaps 33;

QY 12 TEAIFPLSMIAETVLDSASAFDGNKGNFSVRESQEDAGTYYLEKGNVLTENIP-----G 67

DB 268 FVAGEVAGAAATLTVGDTLSSGAGT-----DVLNHWQAAYALPTG-VTISGIETMNV 321

QY 68 TGTATITKSCFNNTRKDLFFETGNGNSLFOYDAG-----TVGAANSSVVDKSTFI 120

DB 322 SGAATTLTNTSSGVTS-LTALNTNTSGAAQYVYAGAGNLTATTAQAANNVAVD----- 374

QY 121 GFSSLSFIASPGSSITTKGAVSCSTGSLTKNVLFSKFNSTDNCGATYAKTSLTG 180

DB 375 GRANTVASTGVTSSTTVGANSASGVTSV-----SVANSSTTTGAIATG 422

QY 181 TTMALFSENTSSKKGAIGTSDALITGNQG--EVSESDNTSDSGAIF--TEASVTI 236

DB 423 GT--AVTYAQGTGNVNTLTQADVTVGNSSTAVTVQTQAATAGATVAGRVGAVTI 480

QY 237 SNNA-----KVSFIDNKVTGA-----SSSTTDMSG-----GALCAVYTSYDT 274

DB 481 TDSAAASATTACKIATVTLGSGAATIDTSALTVNLSGTGSLIGRGALTA----- 533

QY 275 KYTLGNOMLFSNNTSTTAGAGIYVKKLELASGULTFSRNSVNGTAPKG-GAIAIED 333

DB 534 --TPTANTLTLLVNGLTIT--CAITDSEAADG--FTTININGSTASSSTIASLVAD 585

QY 334 SGEISLSDSGDIV-----FLGNTVST-----TPQTNRSSIDLGTSA 371

DB 586 ATTLLNSGARVYTISSHAAALGTITVNSVATLGAEANGLVFTGAGRDSILLGAT 645

QY 372 KMTALRSAGRAIYFDPITTSSTTV-----TDVLKVN----- 405

DB 646 K--AIVMGAG-----DDITVYSSATLGLAGSVNGGDDGVLANVNGSSPSADPAFGF 697

QY 406 ET--PADASLOYTGNI--FTGKLSIETEA----- 432

DB 698 ETLRPAAGAAQSHNANGFTALQCATAGATFTVAVNNGVTVLAATGTTVTLANAT 757

QY 433 ---DSKNITLSKLLQPYTLISGTLSL-----KHGVTLDTQA-----FT 466

DB 758 GTSDVENLT--LSSSALAAGTVALAGVETVINAATDNTTAHVDTLLQATSAKSIYVT 815

QY 467 QQADSRLENDVITTEPADTSTI-----NNLVINSSIDGAKKIEKAKSK 514

DB 816 GNAGINLTNTGTAVTSPDASAVTGTAPAVTFVSANTTGEVVTIRGGAGADSLTGSATA 875

QY 515 NLTLL--SGTILLDPTGFEYFNHSLRNQSYDIELKASGVSTAVTAPPIMEKPHY 571

DB 876 NPTIIGGAAADLVYTGGT---DTFGTGADIDFINAIGSTAFVITTTDAVADKDL 931

QY 572 GYOGTWGPITWGT-GASTT 589

DB 932 VGISTNGAIDAGAFGAAYT 950

RESULT 5

YH8_YEAST STANDARD; PRT: 881 AA.

AC P47033;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE HYPOTHETICAL 89.2 KDA PROTEIN IN SCPL60-SMCS INTERGENIC REGION.

GN YJL078C OR J1027.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;

OC Saccharomycetaceae; Saccharomyces.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE: 96093911.

RA Miosga T., Schaaff-Gerstenschlaeger I., Chaltatzis N., Baur A.,

RA Botes E., Founthier C., Schmitt S., Veltan C., Wilhelm N.,

RA Zimmermann F.K.;

"Sequence analysis of a 33.1 kb fragment from the left arm of

OY	70	TAITSCENNTKGDLTFTGNGNSLTLFQTVAGTVAGAAVNSVDSKSTPIFSSLSFLA	129
Db	737	ANV-----GSLHFSSGTSIVSGTV--GGQGLAKNLNLLDGGTVKFLGDITF--	783
OY	130	SPGSSITTGKAVSCSGTSLTKNVSLLFSKNESTIDNGCAITAKTLLSLGTMSALPSE	189
Db	784	-NGGTRIEGKSIILS--SNYITDHE-----SADNTG-----TLEFVNIDPTV--	825
OY	190	NMSKKGGAIGTSDALITTGQGVSVS-----DNMSDS--GAATF-----	229
Db	826	-TLNKGATFPGYLKQAVSG--PGNAINEIGNGVAAHAIADVSLSFENASLSLFLSCT	883
OY	230	-----TEASVYISN--NAK--VSFIDNKVTGASSTYTGSSGCAICAKTSHDKV	276
Db	884	PLDVLTIKSTVGNGHVDENAPILVSGIDSMIN--NGVYIGQKN--IIALSLSDNSI	939
OY	277	TLTGQMLLFESNNTSTAG-----GAIVYKLELASGCL--TLFSRNSVNG- 321	
Db	940	TVNSMTVLVAGIRTKIKNGVITLISGGIPNNPGVYLGLENGDPKIKQVTFITDVNNGS	999
OY	332	-----TAPKGAALIEDSGELSLASDVG--DIYFLGNTVTSSTPGTNSSIDL 367	
Db	1000	IATATVITINDVITLTGTGIACTDGDGKITLIGSINGNANKAFVDRTFSHT-----SMTV 105	
OY	368	GTSAKMTALRSAAGRAIFYDDPITTGSSITVTVDLVKYNETHPADSALOYGNITFTGEXLS 427	
Db	1054	STKANQ-----GTAVYLGALVAGNIGSSDIPVASRFTGDSGVGLGNI-----HS 1100	
OY	428	EFEADSKNLTKLLOPYTLISGGLSLKHGVTLQTOAFQOADSRLKEMVGTLEPADTS 487	
Db	1101	QNIDEGTYNLT--LNSDVLIGGGTAAINGEIDL-----NNLIFANGSTWGNNTS 1151	
OY	488	---TINNVLINISSIDAKKAKIETKAT-----SKNLTSGTITL----- 524	
Db	1152	LSTTLVNSGNWGVIVAEQAQVWATTTGTTIKIDNMNANSSGQVYTLTLOGGARENG 1211	
OY	525	-----LDPTGT--FYENHSLRP--QSY-----DILELKASGTVSTAVTPDPIMEKEF 569	
Db	1212	TLGAPNEFDVTNNIFVYKELLRDANODYVLTRTNDVLNV--VTTAAGNSAIAAPVHONI 1270	
OY	570	HYGVQGT-----WGEIVWGASTATFENMTKTYIINPERISL 609	
Db	1271	AICLESTDTAAVNMMLLAKOSSDVAFTIGAIADTGA--VAIYNLNDT-----OKTDL 1323	
OY	610	VPNSLMNAFIDISSLAHL--METANES-----LOGDRAF-----WCAGLSNFFHKD 653	
Db	1324	LGNRL-----GALRYLINSSETADVGGSELTGAVSSGDEALIDQVSGYAWKPFYVIAED 1376	
OY	654	STKTRGGRHLSGGVYVIGNLIHT--CSPKILSAFCQLFGHDROYFPAKNOGVYGGTILY 712	
Db	1377	KKGGLAGKAKTATGAVVVG--LDLTLANDNLGATIGITFTD-----IKHOD-----YK 1422	
OY	713	QHNETYISLPCKLRPCSL--VYPELIPVLESNGLSTHTDNDLKTXYTY----- 761	
Db	1423	KGDKT-----DIKLSFSLYGAQQLVKNFPAAGSAILFLTNKVKSSQGYFPFDANGKNK 1476	
OY	762	PTVAGSMGNSDFALEFGGRAPICLDSEALFEQYM--PFMKLQEVYAHQSEFKQGT 815	
Db	1477	QIAAGNYDN-----ITFGNLMFQYDVALQGLVLTMPAGLSYLKSSNENYKXETGT 1527	
RESULT 7			
OMPB_RICPR STANDARD; PRT; 1643 AA.			
AC Q53020: Q9ZCM0: 30-MAY-2000 (Rel. 39, Created)			
DT 30-MAY-2000 (Rel. 39, Last sequence update)			
DT 30-MAY-2000 (Rel. 39, Last annotation update)			
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)			
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)			
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)			
EN (120 KDA OUTER MEMBRANE PROTEIN OMBB); 32 KDA BETA PEPTIDE).			
EN OMBB OR SPA OR RP704.			

Rickettsia prowazekii:
Bacteria; Proteobacteria; alpha subdivision; Rickettiales;
Rickettsiaceae; Rickettsiae; Rickettsia.
(1)
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

STRAIN-BREINL:
MEDLINE: 91045972.
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.:
RT "Characterization of the gene encoding the protective paracrystalline-
RT surface layer protein of Rickettsia prowazekii: presence of a
RT truncated identical homolog in Rickettsia typhi."
RT Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BREINL:
RA Moron C.G., Yu X.J., Walker D.H.:
RT "Sequence analysis of ompB of Rickettsia prowazekii."
RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]

SEQUENCE FROM N.A.

STRAIN-MADRID E:
MEDLINE: 99039499.
RX Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Siccardi-Zentgraf T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.:
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RT Nature 396:133-140(1998).
RN [4]

PARTIAL SEQUENCE.

STRAIN-BREINL:
MEDLINE: 92114896.
RA Ching W.M., Carl M., Dasch G.A.:
RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.";
RT Mol. Immunol. 29:95-105(1992).
RN [5]

IDENTIFICATION OF CLEAVAGE SITE.

MEDLINE: 92104668.
RA Hackstadt T., Messer R., Clepiat W. Jr., Pascock M.G.:
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RT deficient in processing.";
RT Infect. Immun. 60:159-165(1992).

-I- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A TICKETTSIAL
VIOLENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.

-I- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.

-I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and thus statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL: M37647: AAA26390.1; ALT INT.
EMBL: AF161079: AAAP2234.1;
EMBL: AJ255273: CAH15140.1;
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1328
FT VARIANT 257 1643
FT VARIANT 1010 1010
FT VARIANT 1450 1450
FT CONFLICT 178 179
FT CONFLICT 191 201
FT CONFLICT 212 212
FT CONFLICT 313 313
FT

 120 KDA SURFACE-EXPOSED PROTEIN.
 32 KDA BETA PEPTIDE.
 V -> A (IN STRAIN BREINL).
 Y -> D (IN STRAIN BREINL).
 A -> S (IN STRAIN BREINL).
 AA -> VC (IN REF. 1).
 TTGAAPLFLGA -> INSRSSTHLVS (IN REF. 1).
 Q -> L (IN REF. 1).
 Q -> I (IN REF. 1).
 Q -> L (IN REF. 1).
 Q -> L (IN REF. 1).

FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SQ SEQUENCE 1643 AA: 169854 MW: 735FDF392E6346CC CRC64:

Query Match 4.9%; Score 232; DB 1; Length 1643;
 Best Local Similarity 20.9%; Pred. No. 2.5e-05;

Matches 183; Conservative 130; Mismatches 330; Indels 232; Gaps 43;

23 TETVLDSASFQDGNKNGFVSRESQEDAGTTLFKGNVT-----
 748 TSIVSGTVGGQGHKLNNLL-----DNGTIVKRLGDTTNGTKIKGKSLIQSNVYT 802
 62 --LENIPTGT-----AATKSCFNNTKGDLFTT--GN-----89
 803 DHEASADNTLTLEFVNDPTITVLNKGAGYGVLYKQVLIISGPGNIVNEIGNAGIVHIA 862
 90 GNSLFTQTVAG-----TVAGAAVNSVVDKSTTFIFGSSLSFIASPGSSI 135
 863 ANGISFENALGTSFLPSGTPLDVLTIKSTVGNGIVDNNAPVAVVSGIDSMNNQIT 922
 136 TTGKAGVSCSTG--LSLTKNVSLFESKNFSTDNGCAITANTLSLTG-----TTMSALFS 188
 923 GDRKNIITALLSGDINSITVANTLYSGIRTKNNQG-----TVLSGGMPPNPGIYGL 978
 189 ENTSKKGAIGTSDALITINGOG-----EVSFSDNTSSDGAALFT--EASVT--ISN 238
 979 ENSPKIKOVTFITD-----YNNUGSIANNVTINDVYTLTGAGTDPKATITLGSVNG 1034
 239 NAKVSFIDNKVITG-----ASSSTTGDMG--GCAICAKYKSTDT--KVLTLGNOMLFS 287
 1035 NAWRFVDSITFSDPRSMIVATQANKGTVTYLGNALVNSISLDIPVASVFTG-----1087
 288 NNTSTAGCAIYKKLELAGGTLFSRNVNGSTAPKGAIAIEDSGELSLADSCDIV 347
 1088 NDSGAGLOGNIYSONIDFGYNTLILNSNVLG-----GGTTAI--NGEIDL--TNLLI 1138
 348 FLGNVTSTPGTNRSSIDLTGSAKMTALRSAGR--AIFVYDPTTGSSTTVDYKVN 405
 1139 FANGT--STWGDNTSI-----STTLVSSNGINQVVAIEADQVNAITTTGTTIKIO 1187
 406 E-----TPADSAIQ-----YTGNITFTGKLESTEAD--SKNLTSKL 441
 1188 DNANANESGTOAYTLIOGARGFNGTLCAPNFAVYTGNIFFAKYELLIRDQNDVYLTINDV 1247
 442 LQVTLISGGLSKH--GVYLOTOAFQOADSRLKMDVGTLLERPADISTINNVI-----494
 1248 LNVYTTAVGSAIANAGVS-----QNISRCLESTINAAAYNNMLANDPS 1292
 495 NISSIDGAKKAKIETKATSKNL--TISGTTILLDPTGTFEENSLRNFOSYDILELKASG 552
 1293 DVATFPGAIAITDASAANTYTNLNDTQTOPLLNSRLGT--KLSAHEFSDVAG--SATG 1348
 553 TVTSTAVTPPPIGKEKFIHQGTWGPFIWGTGAST--TAFPMNTKTYIYIPNERIGS 608
 1349 AVSS-----GDEAEVSY--GVMAKPFYNAIEODKKGAGIAGKAKTATVVVGLDLAS 1398
 609 LVPNSLMNAFIDISLHYLMETANEGLOGDRATFCAGLSNFKHDKSTKTRGRFRHISGCV 668
 1399 --DNLMGAIGITKTDIKHODYKKGDKT-----INGLSLSLY-----GSGQLKKNF 1444
 669 VIGGNLHTCSDKILSAA--FCOLFGRDRDYFAAKN--QGTVYGGTILYYOHNETYISLPCK 724
 1445 FAQGNMFIETLNKVKSKORFEEFESNGKMSQOIAAGNYDNMTFGGNLIFGVD--YNANPNV 1502
 725 L--RPGSLSVPTPEIPLVLSGNLSYTHDNDLTK 757
 1503 LVTPMAGLSY-----LKSSENENYKETGTTPVANK 1530

RESULT 8
 Y109_YEAST STANDARD; PRT; 995 AA.

P40442;
 01-FEB-1995 (Rel. 31, Created)
 01-FEB-1995 (Rel. 31, Last sequence update)
 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHEICAL 99.7 KDA PROTEIN IN SDLI 5'REGION PRECURSOR.
 GN YIL169C OR Y19402.07C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 Saccharomycetaceae; Saccharomyces.
 RP [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 Raulier C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 Walsh S.V., Whitehead S.;
 Submitted (DEC-1994) to the EMBL/GenBank/DDJ databases.

CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; 246921; CAA87023.1;
 DR SGD; S0001431; YIL169C.
 KW Hypothetical protein; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 995
 FT CARBOHYD 26 28
 FT CARBOHYD 35 35
 FT CARBOHYD 468 468
 FT CARBOHYD 664 664
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 995 AA: 99735 MW: F63E287A03F137EC CRC64;

Query Match 4.8%; Score 230; DB 1; Length 995;
 Best Local Similarity 21.2%; Pred. No. 1.6e-05;
 Matches 180; Conservative 136; Mismatches 355; Indels 180; Gaps 35;

11 STFAIFPLMIATETVLDSASFQDGNKNGFVSRESQEDAGTTLFKGNVTLENIQGT 70
 10 AALALVYSOALGOYYSNSTISSNSTRVSVSS-----SGSVSI--SSSI 55
 71 AATKSCFNNTKGDLFTTNGNSLLFQTVAGVAGAANSVVDKSTTFIFGSSLSFIAS 130
 56 AETSSATDILSITQASASTSGVSSV--GPRSSSVSVSSVQS-----SSVSVSS 107
 131 PGSSITTGKAGVSCSTG--SLSLTKNVSFLFSKNTFDNGCAITAKLSLTGTTMALFSE 189
 108 SVSQSSSASDVSSVSQASSTSDVSSVSQSSASDVSSVSQSSASDVSSVSQ 167
 190 NTSKKGAIGTSDALITIGN-----QGEVSFSDNTSSDGAALFTFEASV 234
 168 SASASADVSSVSQASSTSDVSSVSQSSASDVSSVSQSSASDVSSVSQ 227
 235 TISNNAKVSFIDNKVAGASSSTTGDMG--GCAICAYKSTDTKRVTLTG-----280
 228 TSDVSSVSQASSTSGVSSGOSVSSASGSSSPQSTASATSGATSNLSLITS 287
 281 --NQMLFNNSTTAGCAIYKKLELAGGTLFSRNVNGSTAPKGA--IAIEDSGEL 337
 288 SASASATASNSLSSSDGTYLPTTTI--SGDLTLGKVATEGVVAAAKLTLLDGKRY 346
 338 SLSDA--SGDIYFLGNTVSTTPTG--NRSSIDLTGSAKMTALRSAGRA--IYFDPT 391
 347 SFSADLKVYGD--LVKKRKETIYPTGEFIDISGENPDYTNFNABEASATSIYSTPS 404
 392 TGSSTVTVDVL-----KVNTPADSALOYT--GNITFTGKLESTEADSK-----435

```

Db 405 FDNSGDLSLSKSKGEVTFSPYNSGAFSPSNALINGSVSGIQRDDTEGVNNGEI 464
Oy 436 NL-----TKLQAPVPLSG-----GTLSLKHGVLTQTAFTQOASRLKLEMDVGTLLPAD 485
Db 465 NLDNGSTVYIEVPVSGKGVNIIISGNLYLHPDTGTQVFERKGVGLAVD-----PTE 518
Oy 486 TSTINNLVINISIDGAKAKIETKATSKNLTSLGTLTLLDPT-----GTFYENH 535
Db 519 T---NATPIPVGVGTGNKIATADITA--LSYDGTGVLTTQGNRQSFALIGFSS- 572
Oy 536 SLRNPQSDILE-----LKAAGTVSTAIVPDDPIMGKEPHYGCGTWGPIVWGTG 585
Db 573 -----SDFSVSGEIPAGAYAYLYLNVNGVATSAASSSTASGASAVTGSTSFASVGTST 627
Oy 586 ASTATPMTKTYGILPNERIGSLVPSNLNMFIDISSLHYMEANBELQDRAPWAG 645
Db 628 AST--SFGASVTGSTASTS--FGASVTGS---TSVYTTLLDYVNAITVYVS-----CS- 674
Oy 646 LSNFPHKDSKTTRGFRHLGGGVYIGNLHT-----GSDKLLSAFCOLFGDRDYEVA 699
Db 675 -----ETTDG-----NGNVYITTTVPCSTTATITSCDETG---CHVS 710
Oy 700 KNOGVYGGTLYQHNEYIISLPCKLRPCSLSVPEIPLVFG-----NLSYH- 749
Db 711 TSTGAVTEIYVSXSKYTTATVTHCDNNGCNTKTYVSECKETSATTAPEKSYTTVTHC 770
Oy 750 TDNDLKTYYT 760
Db 771 DDNGCNTKTYT 781

```

RESULT 9

```

YEAL_ECOLI
ID YEAL_ECOLI STANDARD. PRT. 1250 AA
AC P45508; P45507; P45506; P39441; P76468; P77487;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 131.2 KDA PROTEIN IN UBIC-NRDA INNERGENIC REGION
DE PRECURSOR.
CN YEAL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426517.
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.:
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RA Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomoto K., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.:
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE: 84272624.
RA Carlson J., Fuchs J.A., Messing J.:
RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]

```

```

RP SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN-K12 / EMG2;
RA Estep P., O'Keefe T., Robison K., Church G.M.:
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN-0V6;
RX MEDLINE: 88201664.
RA Hussain K., Elliott E.J., Saimond G.P.C.:
RT "The complete sequence of gyrA."
RL Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION.
RX MEDLINE: 96032851.
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Darchin A.:
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes."
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -1- SIMILARITY: TO E.COLI YDER.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC MANY FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AEO00313; AAC75293.1; -
DR EMBL: D90855; CAB21981.1; ALT_INIT.
DR EMBL: D90854; CAB21978.1; ALT_INIT.
DR EMBL: K02672; -; NOT_ANNOTATED_CDS.
DR EMBL: U30459; AAA74094.1; -
DR EMBL: Y00544; -; NOT_ANNOTATED_CDS.
DR EMBL: EGI2850; YEAL.
KW Hypothetical protein; Repeat; Signal.
FT SIGNAL 1 23
FT CHAIN 24 1250
FT DOMAIN 919 948
FT FT 28 30
FT CONFLICT 40 40
FT FT 65 66
FT CONFLICT 431 431
FT FT 433 434
FT CONFLICT 478 478
FT CONFLICT 773 773
FT CONFLICT 853 853
FT FT 923 924
FT CONFLICT 948 994
FT SEQUENCE 1250 AA; 131152 MW; 1/F98C05E299FC95 CRC64;

```

Query Match 4.8%; Score 230; DB 1; Length 1250;

Best Local Similarity 24.1%; Pred. No. 2.2e-05; Mismatches 210; Indels 130; Gaps 28;

Matches 134; Conservative 83;

```

Oy 102 TVAAANVSVVDSTFFIGFSSLFASPOSSITTKGAVSCSTGSLTKNYSLFK 161
Db 49 SLSTGTDWSTLADQ--WLVSDMTNNA--GGAVFLOQGA---EFSLLPENETGMTLFAN 102
Oy 162 NFST--DNGCAITAK--TSLTGTMSALFSENTSSKKGAIGTSDALTTGNGEVS 215
Db 103 NTVAGEVNNCAIFAKENSTLNT---DVFEAGNVAGVGAYISSG---TNDTGAVD 154
Oy 216 -----FSDTSSD--SGAIFTEASVTISNAKVS---FIDNKVYTAASSSTTDDMGGA 264
Db 155 LRVTNAMFRNNIANDKGGAITY-----TINDVYLSVIDPNNQAYTSTYSGD--GGA 207

```

```
QY 265 ICAYKSTDPK-----VTIGNOMLFSNNSTAGCAIYVKLEL-----AS 307
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 208 IDVTNNSSGKHPGTYIVNN--TAFTNNTAEYGGAIIYNSVAPYLDISVDDSYSON 265
QY 308 GGLTFRSSVNG---GTAPKGAIAIEDSGELSLADSDIVELGNTVSTPGNRRS 364
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 266 GGLVDENNNAAGYGGPSSAAGFMVLGLSEVTFPDADOKTLVIGNT-----ENDGA 318
QY 365 IDIGTAKMTALRSAGRAIYFDPTTGSSTVVDLVKNETPADSALOYTNI1FTGE 424
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 319 VD---SIAGTGLITKTGSG---DLVLAANNNDFTGEMQJEN--GEVTLGRSNLSMNVGD 369
QY 425 KLSFEADSKNLTSLKLOPYTLISGGLSKHGVTLQT--QAFTOQADSLNDVETTL 482
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 370 THCODDPDQYGLT---IGSIDOYONOAELNVGSTOQTFVHALTGFGONTLNDAG--- 422
QY 483 PADTSTINNLVINISS---IDGAKRAKIETK-----ATSKNLTSLGTLILDPDTFFE 533
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 423 -----GANTVNGSGFAGIEGAGOLITINQNGSYVLAKQSMALGDIYVDGAYLSLE 475
QY 534 NMSL-----RNPQSY-----DILELK-ASGVTVSTAVTPDPI 564
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 476 GDADADLTALDDPOSIVLNGVLDSLDFSTWQSGTSYNDGLEVSQSGGVIGSDVVDLA 535
QY 565 MGEKPHYXOGTGWPIY 581
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 536 GGDNLHIGDGDKGVY 552

RESULT 10
FIG2_YEAST STANDARD: PRT: 1609 AA.
ID FIG2_YEAST PRT: 1609 AA.
AC P25653:
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE FACTOR INDUCED GENE 2.
GN FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 92397594.
RA Wilson C, Grisanti P, Frontali L;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
RT cerevisiae chromosome III contains two new open reading frames."
RL Yeast 8:569-575(1992).
CC - FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC - INDUCTION: BY MATING PHEROMONES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X59720; CAA42254.1; -.
DR PIR: S19504; S19504.
DR PIR: S25345; S25345.
DR SGD: S0000685; FIG2.
SQ SEQUENCE 1609 AA; 166049 MW; 7D66AD7F85A7B852 CRC64;
```

Query Match 4.7%; Score 226.5; DB 1; Length 1609;
Best Local Similarity 21.1%; Pred. No. 4.8e-05;
Matches 206; Conservative 130; Mismatches 363; Indels 279; Gaps 43;

```
QY 11 SPTAIPPLMI---ATETVLDSSASDGNKNGNFVSRESEDDGCTYTLRKGNVTLENNING 67
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 362 AITSLDPISSVGSTASSVGISTANFSTQGNSTV--VPESTASSGSOYQDMSSSSPLPLSQT 420
```

```
QY 68 TGTATTKSCFNNTKGDLEITFGNGN--SLLFQTVDAQTAVGAANNSSVYDKSTTFIGFSSL 125
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 421 TWVIVNTT---WTQGSVSTSPAVYSTATKTVD--GVITEVYTWCPLOTKQSOAGVSS- 475
QY 126 SFIASGSSITTKGKAVSCSTGSLTKNV-----SLLFKNF----- 163
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 476 S1SSYPQASSFSGSS1LSSNSSTLAAANNVPESTAGSSGOYODWSSSS1PLSQTWVWIN 535
QY 164 STDNGCAITAKT-----LSLTGTMSALFSENT 191
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 536 TINTQSVSTSPAVSNATKTVGVITEVYTWCPLOTKQSOAGVSSITSAQTSP 595
QY 192 SS--KKG-CAIDTSDALITGNO-----GEV---- 214
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 596 SSILTLGISTLQLSDA--TEKGTETINTHLMTESTSITEPEFGTSIDSEFLCTSEVNLAS 654
QY 215 ---SESDNTSSDGAALFEASVTTISNNK-----VSFTDNKVTGASSST 256
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 655 SLSS1PNSFSSSGSDRTITNSTVTFGSTSKYPTSVSNPTEASOHVSSSVNSLTDFTSNS 714
QY 257 TGDMSGCAICA--YKTS-----TDYKVTLNGOMLFSNNTSTAGCA----- 297
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 715 TETI--AVISNIMHKTSNKDYSLFTTQKTSQKQTLVSTVTTVNGAATYTTWCPAS 771
QY 298 -----IYKKLELASGL-----TLFSRNSVNGTAPRGALAIEDSGELSLADS 343
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 772 SIAYTTSISYKTLVLTVEYCSHSECTPYITSTVATSTIP--LLSTSSSTVLSSTVSE 828
QY 344 G-----DIVFGNTVSTSTPGTNRSSIDLGTSAKMTALRSAGRAIYEDDITGGSSPT 397
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 829 GAKNPASAEVTTINTQVSATSEATSTQVSATSAATABESS-----TISQYST 877
QY 398 VYDVLKNET-----PADSALOYTGNIIFTEKELSTEADSKNLTSLKLOPYT 446
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 878 ASEITSTLQTFHTTGSLLFPALSTEMINTVVSARKTLISTEVCSHSCVPIVTEVY 937
QY 447 LSGGTSLKHK--GVTLQTOAFQOQADSRLEMDVGTLEPADTSTINNLVY----- 494
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 938 TSKGTPSNHSSQTLQTEVEVTLSSHQTVINST--EVCNSICIPYITTSVQMSSTPPT 995
QY 495 -----NISSIDAKKAKIETKATSKNLTSLGTLILDPDTGFYENHSLRNPQSYD--IL 546
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 996 YLTSSSTSSSLASTKSSLE--ASSEMSFTSVSTQSLPLAFCSKRSRTSVSQSMNVL 1053
QY 547 E---LKASGTVSTAVTPDPIGGE--KEHYGIQ--GTWGPYWGTAATATFNMKTGCG- 599
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1054 TMTIMSSSNVISTNEKPESTTSPYFSSGYSLPSSSTPSQYSLSTATT--TINGIKTYVT 1112
QY 600 --IPNER-----IGSLVPSNLMAFIDISSLHYMETAN--EGLQGDRAFW 642
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1113 TWCPLAEKSTVAASSQSSSVRFVSSSKPSSLSQTSIQYTLSTATTITSLKLVYTTW 1172
QY 643 CAGLSNFFHKDKSTRGRFRLHLSGGYVIGCNLHTCSD--KILSAFC-----OLF 690
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1173 C-----LTSKTLGATQTSSTAKRITSASATSTISLSTSE 1213
QY 691 GNDROYFAKNGCYVYGGTLYIQHNHETIYSLPCKLRPGSLVPPETIPLFSGNLSY--TH 749
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1214 SESSSGYLSKG---VCSGTECTQDVPTQSSSPAS---TLAYSPS---VSTSSSSSFTT 1263
QY 750 TNDLTKYTYTPVTKGS 767
| : : : : | : : : : | : : : : | : : : : | : : : : |
Db 1264 TASTLSTHTSVPLPSS 1281
```

RESULT 11
YDBA_ECOLI STANDARD: PRT: 2003 AA.
ID YDBA_ECOLI P76087; P76088; P76856; P76857; P76859;
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 205.9 KDA PROTEIN IN PAAY-ACPD INTERGENIC REGION.
 OS YDBA OR B1401/B1405.
 GN Escherichia coli.
 CC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Pena N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.?
 RA "The complete genome sequence of Escherichia coli K-12."
 RT Science 277:1453-1474(1997).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 97251357.
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
 Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.,
 RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map."
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 464-2003 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 92190338.
 RA Moszer I., Glaser P., Danchin A.;
 RT "Multiple IS insertion sequences near the replication terminus in
 RT Escherichia coli K-12".
 RL Biochimie 73:1361-1374(1991).
 CC -I- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).
 CC -I- CAUTION: THIS IS A CONCEPTUAL TRANSLATION: THE GENE CODING FOR
 CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2/D1S30C ELEMENT
 CC BETWEEN AMINO ACIDS 839 AND 840.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000237; AAC74483.1; ALT_SEO.
 DR EMBL: AE000237; AAC74487.1; ALT_SEO.
 DR EMBL: D90778; BAA15009.1; ALT_SEO.
 DR EMBL: D90778; BAA18880.1; ALT_SEO.
 DR EMBL: D90779; BAA18881.1; ALT_SEO.
 DR EMBL: X62680; -; NOT_ANNOTATED_CDS.
 DR ECGENE: Egl1307; YDBA.
 KW Hypothetical protein.
 FT CONFLICT 489 I -> V (IN REF. 2).
 FT CONFLICT 495 I -> V (IN REF. 2).
 SQ SEQUENCE 2003 AA: 205949 MW: B83A12C8B53220EE CRC64;

Query Match 4.7%; Score 224.5; DB 1; Length 2003;
 Best Local Similarity 20.0%; Pred. No. 8.5e-05;
 Matches 168; Conservative 130; Mismatches 281; Indels 251; Gaps 39;

QY 28 DSSASFDGNKNGNFSVRESQDACTTYFKGNVTLEN---IPGTGTAITKSCFNNTKGD 84
 DB 702 DNNVTLLDGKLT---VVSDEVTYSRQSNLFDGSAKRTSLGVIGDGNVY-----NMNG 752
 QY 85 TFGNGNSILFQYTDAGVAGAAVNSSVY-----DKTFTTIGSSLSF-----127
 DB 753 ELIEKNAALADGSOVTLRTGYSTVIVSGESSVYLNGDTTISGEPPLGFAGVIRVOD 812

QY 128 -----IASPGSITTKGANVSCSTGSLTKN-----154
 DB 813 KALLEIGSATLLMODIDSEFHHGTRTVEIONLGFAYVTEGTTGINSGLTSLONKDP 872
 QY 155 -----VSLFSKNEFTDNGAITAKTL-----LTGTTMSALFSENTSSKKGAITYTS 202
 DB 873 APSPIVLLATIGSATNAGITTGATVEQHSVFNKYSIGTINSFFFNNDVSSITGLVAQSN 932
 QY 203 DALIIT-----GNQEVFSFSDNTSSDGA-----AIFTEASVT-----ISSNAKVS 243
 DB 933 STIINTSGIIDLGRGSVGMALADSTAEVQKRTLDMSWVDANDTTAMRDIASNAID 992
 QY 244 FIDKVVYGAASSTGDMGGAICAKRTSTDTKYTLTGNOMLFPNNSTAGAAIYKKL 303
 DB 993 FGTGVGVTDSYSGAKNATY-----NQL-----GGVITLYN- 1025
 QY 304 ELASGGLTLF--SRNSYNGTAPKGAIAIEDSGEL--SLADS--GDIYFLGTVTSTT 357
 DB 1026 --AGAGMAVGAASVTINQGT-----INLEKNYNDISLAANTVGMAYVEHGTAINDQ 1077
 QY 358 PGTNRSSIDLTSAKMTALSAAGRAIYFPDPTTGSSTVT-----DYLKVN 406
 DB 1078 TGV--ININVT-----GOAFY-----NDGTGITVNTGTTCTPGVCOSGNEYNN 1119
 QY 407 TPADSALOYTG--NIIFTEGLKS--ETEAADSKNLTSKLQPVLTSGGLTLKKGVTLOQ 463
 DB 1120 TDDFTSLITYGDTTTSGETVITNKAAYTDKLAGVAVNSGLISGQITVSSGLENTS 1179
 QY 464 AFTQOADSRLMDVGTLLPADTSTINNLVINISIDGA---KKAIEKATSKNLTLS 519
 DB 1180 GGI--INNLVLDKGAVIKNAGVMT--NNVDVSGILNNAEMTAQITMNAAGDSLIVNT 1236
 QY 520 GTIT-LDDPTGTFEENSL--RNQSYDIIELEKASGV-----TSAVYPPDPIMGKFX 570
 DB 1237 GTIKKIYQNVAGVFNNSSVYGRMMSAGVFNNQTDGAIMGAALTGAVANN-----1288
 QY 571 YGYGTGMPVIWVGAST-----TATFN-----WTKTGY--I 600
 DB 1289 ---EGTNINLSSSSSGNNTGMLLEVYNNNSAFNNRGEFILDNKNVHINOSGLTYNTGHMT 1345
 QY 601 PNERIGSLVPSNLIW-----NAFIDISSLYHMETANEGLOGRAFWCA--GLSNFP 650
 DB 1346 SNSHNGAV---NMWGGNGRFINDTIDVSAKSIVASANNAGDQ--NAFVANNQNGVINF 1401
 QY 651 HKDSTKTRRGFRHISGVYVIGGNHTCSDKILSAFQLEFRGRDVFVANKQTYVGGTL 710
 DB 1402 HDSASAVK-----VTHSNFIAQNDGIMINI SGTGAVAMEGD---KNAOLVNNGTI 1447

RESULT 12
 OMPB_RICJA STANDARD; PRT: 1656 AA.
 ID OMPB_RICJA
 AC 006653;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
 DE (CONSTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
 GN OMPB.
 OS Rickettsia japonica.
 OC Bacteria: Proteobacteria; alpha subdivision: Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein romp B of Rickettsia
 RT japonica".
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR

CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIOLENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB003681; BAA20138.1; -
CC Antigen: S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
CC CHAIN 1 1338 32 KDA BETA PEPTIDE.
CC FT CHAIN 1 1338 1656
CC DOMAIN 528 533 POLY-GLY.
CC SEQUENCE 1656 AA: 168097 MW: 31326969CD5999F CRC64;

Query Match 4.7% Score 223.5; DB 1; Length 1656;
Best Local Similarity 20.0%; Pred. No. 7.4e-05;
Matches 186; Conservative 113; Mismatches 336; Indels 295; Gaps 44;
QY 22 ATETVLDSSASFQDNKNGFVSRESDACTVYLFKGNVLE-NIPQTAITKSCFNN 80
DB 386 ASIVATIQNSFGTDEGTLNLAQVTPD--TMTLTGNTFGDANNPQ-----NT 431
QY 81 KQDLTETGNGNSLLFQVVDAGTGAAGAVNSVVDKSTTFGFSLS-----FIASPS 133
DB 432 AGVITFAANT-----LASASADANVAVTNNITAIKASGVVQLSGTHAEIRLGAGS 486
QY 134 -----SITGK-----GAVSCSTGSLSTKXVSLSKSKNSTDNGG---A 170
DB 487 VEKLADGTIVNGVQTVLVGVLAAQIILD-GSATITTDIG-----NGGGGA 535
QY 171 ITAKTSLTGTTMSALFSENTSSKKGGAID--TSDALITGNGEVSF-----SDNT 220
DB 536 LOSITLANDTKTLTGLGANNIISANGTINFGANGTIKLTSTQNNIVYDCDLAIADQ 595
QY 221 SSSGGAIFFEASVTISNNKVSFIQNKVTG-----ASSSTGMSGGALC----- 266
DB 596 GAVDASSLINAQILITISGTLIGANNITLIGQFNIGSSKTT--LNGGVAINELVIGNNG 653
QY 267 ---AKTSTDTKVT-LTGNOMLLFS---NNTSTAGAIYVKLELASGGLTLEFSNSV 318
DB 654 SVGFANHTVLTFTTNAAGCKIIFNPVNNNTTLAGT-----NLSGANLAEINFG 707
QY 319 NGSTARK-----GGAIAIEDSGELISADSGIVFL--GNTVSTTPG-----TNBS 365
DB 708 SKARADTVLVNVEGVNLVATNITTTDANGSFVFNAGGNIVSGTGGQGNKFTVAL 767
QY 366 DLCTSKMFLRSAGRAIFYDPTTGSSTVTDVLKVNTPADSLAYTQNIIFGEK 425
DB 768 DNGITVK-----FLGNATPFGNTTIA-----ANSLDQJSGN--YTADF 803
QY 426 LSETEA-----DSKULTSKLQPYTLG-----GTLSEKHGVTLQ 461
DB 804 IASADGTVIEFVNTGPIVNTLKNQAVPVNALKQITVSGGNVNVNEISAGNVHGAMTD 863
QY 462 TQAFDQADSRLEM-----DVGITLLEPADSTINNLV-----INSSID----- 500
DB 864 TIAFENSSIGAVLFLPSGIFPNAGNTIPLTIKSTVGENETAEFSPSVYSGVDVSIAD 923
QY 501 ---GAKKAKIETKATSKNLTLSGTTLLDPTGTGFENHSLRNPQSDILELKASGTVTS 556
DB 924 GGVIGQNNITVGLGSDNGIIVNATTLVAGIGTINN-----OGVTVL 967
QY 557 TAVTPDPIKGEKHYGQGTWGPVWGCGASTATFNMTKGYIIPPERIGSLVPSLWN 616

DB 968 SGGVPT-----PQTVGLGTGIGASKFQVTF-TDY-----NNLGNIAITN--- 1009
QY 617 AFIDISLHYMETANEGLO-----GDRAFWAGL-----SNFHKD----- 653
DB 1010 -----TTINDGTVTTGGTGAAGGIAGTDFDGTITLGSVNGNANVPADGISN 1057
QY 654 -----STYTRRG-FRHLSGYYVIGNLTGCSDKILSAFCOLFGFRDYPVAKNGTVY 706
DB 1058 STSMVTTTANNGTVYILGNAFV--GNIDSDTPVASVR-----FTGSNNAGL 1104
QY 707 GGLTYQHNE--TYSLPKLRPCSLSTYPTETPULFSGNSLSTHTDNDLTKRYTTPPV 764
DB 1105 KGIYISQVVIDFGTY-----NLGIYNSN--VILGSGTAINKIDLLNTLTFPAG 1152
QY 765 KSGWGN-----SFALEFGRAPICDESA 789
DB 1153 TSTWGNNTSIETTLTLANGNIGHIYAEGA 1182

RESULT 13
EGT2_YEAST STANDARD; PRT; 1041 AA.
ID EGT2_YEAST
AC P42835;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE EGT2 PROTEIN PRECURSOR (EARLY G1 TRANSCRIPT 2).
GN EGT2 OR YNL327W OR NO320.
OC Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1676;
RX MEDLINE; 95373280.
RA Maftah M., Nicaud J.-M., Levesque H., Gallardin C.;
RT "Sequencing analysis of a 15.4 kb fragment of yeast chromosome XIV
RT identifies the RPD3, PAS8 and KRE1 loci, five new open reading
RT frames.";
RL Yeast 11:567-572(1995).
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RX MEDLINE; 96251274.
RA Kovacech B., Nasmyth K., Schuster T.;
RT "Eg2 gene transcription is induced predominantly by Swi5 in early
RT G1.";
RL Mol. Cell. Biol. 16:3264-3274(1996).
CC -1- FUNCTION: SEEMS TO BE INVOLVED IN THE CORRECT TIMING OF CELL
CC SEPARATION AFTER CYTOKINESIS, AS SEPARATION OF MUTANT DAUGHTER
CC CELLS IS DELAYED. COULD EITHER BE AN ENZYME NECESSARY FOR GLUCANS-
CC DEGRADATION OF THE CELL WALL AT THE NECK REGION BETWEEN MOTHER AND
CC DAUGHTER CELLS OR A REGULATORY PROTEIN CONTROLLING THIS METABOLIC
CC STEP. EXCLUSIVELY EXPRESSED BETWEEN THE END OF MITOSIS AND EARLY
CC G1; INACTIVATED BEFORE CELLS PASS START.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z46259; CA86371.1; -
CC EMBL: Z71603; CA96259.1; -
DR SGD: S0005271. EGT2.
KW Glycoprotein; Repeat; signal; Cell cycle; Cell division; Mitosis.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1041 EGT2 PROTEIN.
FT DOMAIN 200 203 POLY-SER.
FT DOMAIN 381 384 POLY-SER.
FT DOMAIN 388 395 POLY-SER.

```

FT DOMAIN 490 493 POLY-SER.
FT DOMAIN 586 589 POLY-THR.
FT REPEAT 457 492 1-1.
FT REPEAT 577 606 1-2.
FT REPEAT 613 647 1-3.
FT REPEAT 716 745 1-4.
FT REPEAT 773 802 1-5.
FT REPEAT 811 840 1-6.
FT REPEAT 849 886 1-7.
FT REPEAT 887 924 1-8.
FT REPEAT 925 962 1-9.
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 526 526 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 544 544 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 636 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 1041 AA; 108494 MW; 01FECF8BA8744CD CRC64;

```

Query Match 4.7%; Score 222; DB 1; Length 1041;
 Best Local Similarity 23.2%; Pred. No. 4.7e-05;
 Matches 163; Conservative 99; Mismatches 280; Indels 162; Gaps 31;

```

QY 4 SPKPFSTFALPPLMATE-----VLDSSAFPGKNGNFSVREOE 48
DB 395 SAPSSNSTFTTPSSLSATEYISISSASISVTQASIDNSTTAVTOSTIAVSSAE 454
QY 49 DAGTYLFGKGNVTLENIPGTAITKSCFNNTKGDLLFTGNGSLLEQTVDA-----GTV 103
DB 455 KLSSTLSTYNSVTIVSSATQHTTTPSYNSP-----TLSSSVLESVIAPPYANTV 509
QY 104 AGA-----AVNSVVDKSTTFIFGSSLFIA--SPGSSITTGGAVSCCTGSLTKNVS 156
DB 510 SGASASASQSTNPYPVNSST-----SSATQLATIAPPAIMITGISISSITNTSVSSTTS 565
QY 157 LFSKNSFTDNGCAITAKTILSLTGTMALFSENTSSKKGAIQTSDALITIGNOGEVSF 216
DB 566 SLSSGFFVNT--AVASGYIILTTTESAQLT-----IGSLIPITITTTTSS----- 614
QY 217 SDNTSSDSGAJTEASVATISNNAKVS---FIDNKVTGASS-----TTGDMSCGAICAY 268
DB 615 TDKTGSNK--VASSTELIAQSVNNSSLSVSTINTNATAAANARNAFTHAHSGSLQSPY 673
QY 269 KIST-----DTVTLTGOMLFLSNMTSTTAGAIYVKKLELASGULLFNSNNGCT 322
DB 674 HSSSLSTIDTKVT-----TATSTISRDGS-----SSLAFTGLNQSVVSTGT 716
QY 323 APKGAIATJEDSGELSLSDSGDIVLGNTVSTPGTNR--SSIDGTSAKMT--ALRSAA 380
DB 717 -DKSDTYVSTSTESAOVFEYDLSLPISTLKPIVVTGTSNNSFFSVSSTKLEATATDK 775
QY 381 GRAIYFDPITTGSTTVTD-----VLKNE-----TPADSALQY- 415
DB 776 GDA---YSVIASSQSAOVLEYGSMPLSTLEPTVIMSTDESGYFTLTCESGQATEYG 832
QY 416 -----TGNIT--FTGEVL-----SETEADSKNLTSKILOVTLSCGTLKRGVTLQ 461
DB 833 SLIPISLDSGVITFTGESVVGSTYVGAOYAOHTSLVPVSTIKGKSTL----- 885

```

```

QY 462 TQAFTHQADSRLEMDVGTLEPADFTSTNNLVINISSIDGAKKAKIEPKATSKNLTLSGT 521
DB 886 -----STEEVVAAGYSTTGAOYAOHTSLV--PVSTIKSK-----TSLSTEEVWAG 933
QY 522 ITLLDPTGFYENHSLRNPOSYDILELKA--SGTVSTAVTPDPPEKGFHYGQGTWP 579
DB 934 STYVD--SAQAEHT--NLVADITLKTSTFOKATIEVCYCTALSSPHATLDAQTIS 989
QY 580 IYWGTCASTATFNTWTKTYGPNE-----RIGSLV 610
DB 990 LPTSSSTLSLTIITWSSSTIKPSPISFGAGAGQLTRIGSL 1033

RESULT 14
AGAL YEAST STANDARD; PRT; 725 AA.
AC P32323;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR.
GN AGAL OR YNR044W OR N3431.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91304412.
RA Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;
RT "The AGAL product is involved in cell surface attachment of the
RT Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin.";
RL Mol. Cell. Biol. 11:4196-4206(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Poll T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC SUBUNIT. S. CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC AGGREGATION DURING MATING.
CC -1- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC A CORE SUBUNIT.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M60590; AAA34382.1; -
CC EMBL; Z71659; CAA96325.1; -
CC DR PIR; S17031; S17031.
CC DR PIR; A41258; A41258.
CC DR SGD; S0005327; AGAL.
CC KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
CC KM Phenomone response.
CC SIGNAL 1 22
CC CHAIN 23 725
CC FT DOMAIN 53 493
CC FT REPEAT 53 149
CC FT REPEAT 395 493
CC FT DOMAIN 182 307
CC FT REPEAT 182 188
CC FT REPEAT 189 195
CC FT REPEAT 196 202
CC FT REPEAT 203 209

```


Page 13

HYPOTHEORETICAL 113.1 KDA PROTEIN IN PRE5-FET4 INTERGENIC REGION.
 GN YMR317W OR YMR924.09.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:5288C / AB972;
 RA Churcher C.M., Louis E.J., Barrell B.G., Rajadream M.A., Walsh S.V.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).
 CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL: Z54141; CAA90835.1; -
 DR SGD: S0004936; YMR317W.
 KW Hypothetical protein; Repeat.
 SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

 Query Match 4.6%; Score 218.5; DB 1; Length 1140;
 Best Local Similarity 1.2%; Pred. No. 8.4e-05;
 Matches 139; Conservative 107; Mismatches 286; Indels 119; Gaps 20;

 QY 18 LSMATET-----VLDSSAFDGNKNGNFSVRSQEDAGTTLTKGNVTLENIPGTGA 71
 Db 129 LSSSTTETETSISSAIGTSPQTSNNCGGS-----SSEPLKSSVLETTASSSDT---TA 181
 QY 72 ITRKCFNNTKDLEFTGNGSLFPQYDAGNVACA--VMSVYDKSTF----- 119
 Db 182 VTSSTF-TLTLDVSSPKRISSGSAVNTSVGTTSDASKVFSSSTSDVSSLSSTSPSS 240
 QY 120 -----IGFSS--LSFIASPGSSITTGKAVSCSTGSLTKN-----VSLFSKNF 163
 Db 241 TISETLPEFSSITLTSSPVSSEAPSATSSVSSEASSTSSVSSEAPLATSSVSSEA 300
 QY 164 STDNGCAITATLTLTGTMTALFSEMTSSKKG-----GAIQSDALITTINGGEVSF 216
 Db 301 PSSSTVSSEAPSTSSVSSEISSTSSVSSEAPLATSSVSSEAPSTSSVSSEI 360
 QY 217 SDNTSSDSGAIFTEASVTTISNNAKVSPFIDKRVGASSSTTGDSGCAICAVKTSIDRV 276
 Db 361 SSTSSVSSEAPLATSSVSSEAPSTSSVSSEAPSTSSVSSEAPSTSSVSSEI 420
 QY 277 TLTGNQML-----LRSNNT-----STTAGAIYVKKLELASGGLTLFRNSVNGCT 322
 Db 421 SSTSSVSSEVSSEATSLVSSSEAPSAISLASSRLFSKNTSVTLVATEASSVTSSL 480
 QY 323 APRKGAIAIEDSGEISLADGDIY-----FLGNTV-----TSTPTGN 361
 Db 481 RPSEETLANSNIISLSTGNTSVTTTSAASLTGSKVSSNSRMAKTSKTSSTSDLS 540
 QY 362 RSSIDLTSAKMALRSA---AGRAIFYDPITTGSSSTVTVDLKYNETPAD-SALQY 415
 Db 541 KSSVIFENSSVTTPSPASISLTPASPLPVSQWSDITSPSEASSISNLASSAPSDNNSTIA 600
 QY 416 TGNITFGEKLTSEADSKNLTKLQPYTLQSGTSLSKKGVTLQTOAFTQOADSRLQM 475
 Db 601 SASLIVTKTNSVSVSVSSITSSSETTNESNLASSITSLSNKATARSLSTSNKTSASNV 660
 QY 476 DVGT-----TLPEADSTINNLVIN---ISSIDGAKKAKIEKATSKNLTLS--- 519
 Db 661 PTGTFSSMSHTVITPTGFTSSASLAINNTVSVSSSLAGVSFSPRESSPTTSLVTSAP 720
 QY 520 GTITLDPGTGTFVNHSLRP-----QSYDLELKASGYTST 557
 Db 721 SYVSSMTTSAPIFNNSTARSPTASFTTESTTSSISVPL-AGGDVTS 769

Fri Nov 24 13:50:14 2000

us-09-428-122-2.rsp

Page 14

Search completed: November 20, 2000, 11:33:09
Job time: 722 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2000, 11:20:05 ; Search time 20.34 Seconds
(without alignments)
4260.138 Million cell updates/sec

Title: US-09-428-122-2
Perfect score: 4774
Sequence: 1 MKSFPKVFSTFAIFPLSM.....FELRGSSRYNVGLGAKYQF 928

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_PROTOZOA:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3865	81.0	746	2	09RB71 chlamydia p
2	2058	43.1	928	2	09SB62 chlamydia p
3	2000	41.9	914	2	08B163 chlamydia p
4	1982	41.5	928	2	09Z398 chlamydia p
5	1965	41.2	936	2	09Z898 chlamydia p
6	1940	40.6	930	2	09RB66 chlamydia p
7	1936	40.6	930	2	09Z393 chlamydia p
8	1915	40.1	926	2	09Z135 chlamydia p
9	1855	38.9	928	2	08B164 chlamydia p
10	1656.5	34.7	839	2	P77792 chlamydia p
11	1641.5	34.4	847	2	P71132 chlamydia p
12	1604.5	33.6	846	2	P71133 chlamydia p
13	1573	32.9	841	2	09Z3A1 chlamydia p
14	1442.5	30.2	922	2	09Z9G5 chlamydia p
15	1430.5	30.0	922	2	09Z4H9 chlamydia p
16	1383.5	28.0	1407	2	09Z899 chlamydia p
17	1377.5	28.9	973	2	09Z896 chlamydia p
18	1246	26.1	712	2	09RB73 chlamydia p
19	1148	24.0	458	2	09RB65 chlamydia p

20	1120.5	23.5	1013	2	084879 chlamydia p
21	935.5	19.6	649	2	P71134 chlamydia p
22	923	19.3	354	2	09RB69 chlamydia p
23	863	18.1	878	2	084882 chlamydia p
24	821	17.2	494	2	09RB68 chlamydia p
25	792	16.6	427	2	09RB70 chlamydia p
26	768	16.1	177	2	09RB72 chlamydia p
27	711.5	14.9	298	2	09RB67 chlamydia p
28	685.5	14.4	1609	2	09Z6U5 chlamydia p
29	685.5	14.4	1609	2	09RB58 chlamydia p
30	672.5	14.1	978	2	09Z895 chlamydia p
31	667.5	14.0	978	2	09RB63 chlamydia p
32	666	14.0	947	2	09Z813 chlamydia p
33	666	14.0	1723	2	09Z812 chlamydia p
34	666	14.0	1723	2	09RB59 chlamydia p
35	665.5	13.9	946	2	09Z880 chlamydia p
36	665.5	13.9	946	2	09RB60 chlamydia p
37	646	13.5	934	2	09Z882 chlamydia p
38	635	13.3	279	2	09Z5Q4 chlamydia p
39	603.5	12.6	938	2	09Z883 chlamydia p
40	603	12.6	975	2	084417 chlamydia p
41	597	12.5	1016	2	084880 chlamydia p
42	576.5	12.1	1531	2	084818 chlamydia p
43	559	11.7	1751	2	084418 chlamydia p
44	523	11.0	1770	2	084419 chlamydia p
45	513.5	10.8	311	2	09RB64 chlamydia p

ALIGNMENTS

RESULT 1
09RB71 PRELIMINARY: PRT: 746 AA.
ID 09RB71;
AC 09RB71;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
GN PMP-3.2.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J138.
RA Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
RA Takeuchi A., Nishida J., Shibata K., Fujinaga R., Yoneda H.,
RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,
RA Ishii K., Shiba T., Hattori M., Kuhara S.;
RT "Comparison of outer membrane protein genes omp and pmp in the whole
genome sequences of Chlamydia pneumoniae isolates from Japan and US";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL: AB033795; BAA85948.1;
DR EMBL: 746 AA; 80272 MW; 516DE0ED717AF965 CRC64;
SQ SEQUENCE 746 AA; 80272 MW; 516DE0ED717AF965 CRC64;

Query Match 81.0% Score 3865; DB 2; Length 746;
Best Local Similarity 100.0%; Pred. No 1.8e-209; Indels 0; Gaps 0;
Matches 746; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 183 MSALFSESTSSKKGAIQTSALITTTGNGEVSFSDNTSSDGAIFTEASVTSNNAKV 242
DB 1 MSALFSESTSSKKGAIQTSALITTTGNGEVSFSDNTSSDGAIFTEASVTSNNAKV 60
OY 243 SFIDNKVVGASSSTTGDMGGAICAYKSTPTKYLITGNOMLLSSNNSTTAGGAIYVK 302
DB 61 SFIDNKVVGASSSTTGDMGGAICAYKSTPTKYLITGNOMLLSSNNSTTAGGAIYVK 120
OY 303 LELASGGLTFSRNVSNGTAPKGAIAIEDSGELSLSDSGDIYFLGNTVSTPTGTR 362
DB 121 LELASGGLTFSRNVSNGTAPKGAIAIEDSGELSLSDSGDIYFLGNTVSTPTGTR 180
OY 363 SSIDLGTSAKMTALRSAGRAIYFYDPTTGSSTTVTVLAKVNETPADSAQYTGNIIFT 422

```

|||||
Db 181 SSIDLGSAKMTALRSAGRAIYFYDPITGSSSTVTDVLKNETPADSALQYTGNIIFT 240
QY 423 GELSTLEADSDNLSKLLQPTLSGGLSLKHGVTLDQATQOADSLEMDVCTTLE 482
Db 241 GELSTLEADSDNLSKLLQPTLSGGLSLKHGVTLDQATQOADSLEMDVCTTLE 300
QY 483 PADTSTINNLVINISSIDGAKKAKIETKATSKNLTSGTTLLDPGTGFENHSLNPOS 542
Db 301 PADTSTINNLVINISSIDGAKKAKIETKATSKNLTSGTTLLDPGTGFENHSLNPOS 360
QY 543 YDLELKASGTVTSTAATPDPIMGEKFEHYGYGTWGPPIWGTGASTTAIFNMKTGYIPN 602
Db 361 YDLELKASGTVTSTAATPDPIMGEKFEHYGYGTWGPPIWGTGASTTAIFNMKTGYIPN 420
QY 603 PERIGSLVPSLNMNAFIDISSLHYMETANEGLOGDRAPFACALSNPFHKDSTKIRGRF 662
Db 421 PERIGSLVPSLNMNAFIDISSLHYMETANEGLOGDRAPFACALSNPFHKDSTKIRGRF 480
QY 663 HLSGGYVIGGNLHTCSDKILSAFCOLFGRDRDYFAKKNQGTGYGGTLYYQHNETYISLP 722
Db 481 HLSGGYVIGGNLHTCSDKILSAFCOLFGRDRDYFAKKNQGTGYGGTLYYQHNETYISLP 540
QY 723 CKLRPCLSYVPTLEIPVLEFSGNLSTHTNDLTKYTTYPTVKGSGWGSFALEFGGRAP 782
Db 541 CKLRPCLSYVPTLEIPVLEFSGNLSTHTNDLTKYTTYPTVKGSGWGSFALEFGGRAP 600
QY 783 ICIDESALFEQYMPFKLOFYVAHQEGFKQGTAREFGSSRLVNLALPIGTFEDKESOC 842
Db 601 ICIDESALFEQYMPFKLOFYVAHQEGFKQGTAREFGSSRLVNLALPIGTFEDKESOC 660
QY 843 QDATYMLTGTYVDLVRNSNDCTTTLRISGDSMKTGTGNLAROALVLRAGNHCFNSNFE 902
Db 661 QDATYMLTGTYVDLVRNSNDCTTTLRISGDSMKTGTGNLAROALVLRAGNHCFNSNFE 720
QY 903 AFSQSFELRGSSRNYNVDLGAKYQF 928
Db 721 AFSQSFELRGSSRNYNVDLGAKYQF 746

RESULT 2
QY 0956P2 PRELIMINARY: PRT: 928 AA.
AC 0956P2:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE OUTER MEMBRANE PROTEIN 5 PRECURSOR.
GN OMP5.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310:
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkelund S.;
RA "Chlamydia proteins containing the GNAI-repeat belong to a subfamily
RT of autoantigenic pathogenicity factors."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ133034; CAB37071.1; -.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 928 AA; 97229 MW; 0590D5206AIDOE1 CRC64;

```

Query Match 43.1%; Score 2058; DB 2: Length 928;
 Best local Similarity 46.7%; Pred. No. 6, 9e-108;
 Matches 444; Conservative 142; Mismatches 320; Indels 44; Gaps 20;

```

QY 1 MKSFPKFFV-STFAIF---PLMIATETVLDSSASFDGKN-GNFSVRSQDAGTTL 55
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MKSQFSLVSLTACFTSCSTVFAATVNIENIGSDSFGSTGTGTPKNT--TTGIDYT 58

```

```

QY 56 FKNVTLNIPGTGTAITKSCFNNTKGDLTETGNGNSLLFQTVADGTVAGAANSSVYDK 115
| : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 59 LFGDITLQNL-GDSALRLKGFSDTTESLSFAGKGYSLFNLIRKS-SAGEGAL-SVITDK 115
QY 116 STTFIGFSSLSIAPGSSITT--GKAVSCSTGSLSKNLSLFSKNFSTDNCGALTA 173
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 116 NLSLGFSSLTFLAAPSVIITPSSGKAVKCG-GDLTFDNNGTLLFKODYCEENGAIST 174
QY 174 KTLSTLGTMSALFSEPNSS---KKGAITQSDALITIGNOGEVSFSDNTSSDSGAIFT 230
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 175 KNLSTKNGTSGISFEQKSSATGKGAICATGYDIINNTAPLFSNNIAAAGALINS 234
QY 231 EASVTLSNNAKVSFIDNKYTGASSSTTGDMGGAICAVKTSFDTKVTLLTGNOMLFSNNT 290
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 235 TGNCTITGNTSLVFSENVY---ATAG--NGAL----SCDADVTISGNQSVTFSGNQ 283
QY 291 STTAGGAIYKKLELAS--GGITLFSRNSVNGTAPKGAIAIEDSGELSSADGDIVF 348
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 284 AVANGGAIYAKKLLTLAGSGGGGIFSNNIVQGTGNGGALISLLAAGDCSLSAEGDITF 343
QY 349 LGNTVATSTP-GTNSSSIDGSAKMTALRSAGRAIYFYDPITGSSSTVTDVLKNET 407
| : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 344 NGNAIYATTPQTTKRNSIDIGSTAKITNLRAISGHSIFPDITANTAADSTDLNLNKA 403
QY 408 PADSAIQYTGNIIFTEKLESTLEADSKNLTSLQPTLSGGLSLKHGVTLDQATQAFQ 467
| : : | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 404 DAGNSTDYSGIVFSEGEKLEDEKAVADNLSTLKQPTVLTAGNLVLRKGVTLDTKGFQ 463
QY 468 QADSLIEMVGTLEPA-DSTINNLVINISSIDGAKKAKIETKATSKNLTSGTTLLD 526
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 464 TAGSSVIMDAGTTLKASTEEVLTGLSPVDSLGKKVVAASAASKNVALSGTILLD 523
QY 527 PGTGFENHSLNPOSYDLELKASGTVTSTAATPDPIMGEKFEHYGYGTWGPPIYW---- 582
| | | | | | : : : : | : | : | : | : | : | : | : | : | : | : | : | :
Db 524 NQGNAYENDLQKQDFSVQSLAGLATTTDVPAPVYATPTHGYGYGTWG-MTWVDDT 582
QY 583 -GTGASTATFEMTKGYIPNPERIGSLVPSLNMNAFIDISSLHYMETANEGLOGDRAF 641
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 583 ASIPKTKITATLMTVNTGYLPNPEROGPLVPNSLWGSFSDIOAGVIERSAULTLSDROF 642
QY 642 WCAGLSNFFHKDSTKRGRFRLSGGYVIGNLHCSDKILSAFCOLFGRDRDYFAVKN 701
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 643 WAGVAVNFDLQDKGKGRKRYRHKSGGYAIGGAQTCSENLISFACQLFGSKDPLVAKN 702
QY 702 QGTVYGTLYYQHNETYILPCKLRPCLSYVPTLEIPVLEFSGNLSTHTNDLTKY 758
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 703 HTDTYGAAYIYH---ITECGFLICLLDKLPGSMXKPLVLEQGLAYSHVSNLTKY 758
QY 759 TTYPTVKGSGWGSDFALFEGGRAPICLDESALFEQYMPFKLOFYVAHQEGFKQGTAR 818
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 759 TAYPEVKGSGWGNNAFNMMLGASHSYPEYLHCFDYAPYIKLNTLYIRODSFSEKGTGR 818
QY 819 EGGSSRLVNLALPIGTRPKESDCODATYNLTLGTYVDLVRNSNDCTTTLRISGSMKTF 878
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 819 SFDSNLFNLSLPIGVKFEKFSDCNDSTDLTSLVPTLIRNDPCTTALVLSGASWET 878
QY 879 GTNLARQALVLRAGNHCFNSNFEAFSOPFELRGSSRNYNVDLGAKYQF 928
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 879 ANNLAQALQAVRAGSHYAFSPMEVIGQVFEVGRSSRIYNNVDLGKRFQF 928

RESULT 3
QY 086163 PRELIMINARY: PRT: 914 AA.
AC 086163:
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE OUTER MEMBRANE PROTEIN 5 PRECURSOR (FRAGMENT).
GN OMP5.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]

```

Query Match	Best Local Similarity	41.9%	Score 2000:	DB 2:	Length 914
Matches	433;	Conservative 141;	Mismatches 318;	Indels 44;	Gaps 20
QY	1	MKSSEPKVF-STFAIF--PLSMATETVLDSASFQDNKN-GNFSYSESDAQTLYL	55		
Db	1	MKSQSMVLVLSSTLCACFTSCSTVFATAENIPDSQDSTNGTYPKNT--TTGIDY	58		
QY	56	EKGNTLENIPEGTATKSCFNENNRKGDLEFGNGSLFQYVADATVGAANSSVVK	115		
Db	59	LGTDTTLQWL-GDSALILKGCSDTTESLSFGKYSLSFLNKS-SAGCAL-VYTTK	115		
QY	116	STTFIGFSSLSFIASPGSSIT--GKVASCSGSLSLTKNYSLSKNESTDNGAITA	173		
Db	116	NLSLGFSSSLFPLAPSSVITTPSGAVKCG-GDLTFPNNGTILFKQDYCEENGALST	174		
QY	174	KTLSLGTMTKSLFSENTSS---KKGCAIQTSDALITGNOGVSSDNTSSSGAIFT	230		
Db	175	KNLSKNTSGSISFEGKNSATGKKGALCAIGTVDTNNTPATLESNNIAEAGGALIS	234		
QY	231	EASVTSNNNAKYSFIDNKVTGASSPTTGMSGGAICAIVTSTDTKYTLGNOMLFSNT	290		
Db	235	TGNCITGNTSLVFENSTV---ATAG--NGAL-----SGADVTISGNSVTFSGNQ	283		
QY	291	STTAGCAIYKKLELAS--GGLTFSRNSVNGTAPKGAIAIEDSGLSADSDIYF	348		
Db	284	AVANGCAIYAKKLTLASGGGGISFESNNVVOGTTAONGAISTLAGDCSLSAEADITF	343		
QY	349	IGNTSTSTP--GTNNSDITGSAKTAIRASAAGRAIYEDPTTSSSTTVDPVLVNE	407		
Db	344	NGNAIVATTPTTKRNSIDIGSTAKTTLNRAISGHSIFYPDITANTADSDTTLWNKA	403		
QY	408	PADSALOYTGNIIITGKEKLETEADSKNLSKLOPVTLSGGLSLKAGVLTOTAFTO	467		
Db	404	DAGNSTDVSGSIVFSGEKESEDEKADYADNLSLKOPVTLTGNVLKRGVTLDTGCTO	463		
QY	468	QADSLEMDVGTLEPA--DSTINNLVINISSIDAKKKKIEKATSKNLTLSGTTITLD	526		
Db	464	TAGSSVINDAGTTLKASTEEVTLTGSLISIVDSIGBEKKVIAASAKNVALSGPILLD	523		
QY	527	PTGTVEYENSLNPSQSYDLELKASGTSTVATPPRIGKEFHYGYQGTGPIYV----	582		
Db	524	NGGNAYENHDLKQDQDFSVQLSALGATATTDPVAPVATPTHYGYGTGW--MTWVDPT	582		
QY	583	-GTGASTATFNTWYTKGYLPNPERIGSLVPSNLMAINFIDISSLHYMTANBSLGDDRAF	641		
Db	583	ASTPTKATATLMTWNTGYLPNPERQGPLVNSLWGSFSDIOALOGVIESALTLCSDRG	642		
QY	642	WAGASLNFPHKOSTTRBGRFHLISGGYVVGNNLHTSDKLTLSAFCQLGRDDYVANK	701		
Db	643	WAAGVANTFLDKRKKEGKRYKHSKGYAIGCAAOITCSEMLISFAFQDLGSKDPLVANK	702		
QY	702	OGTYVYGGTLYOHNETYISLPCKLRPCLSYVP--TEIPRIFSGNLSTYTHDNDLTKY	758		
Db	703	HDYTAGAGAYIOH---ITECGFGICLLDKLPKGSWHSKRPVLECOLAASHVENDLTKY	758		
QY	759	TTYPTVKGNGNDSFALLEGRCAPICLDESALFEQVMPMKLOFYVAHOGFEQCTEAR	818		
Db	759	TAYPEKSGWGNNAEFMMGLGASHSYPEYLHCDDYATPAIKILNLTIRDSPEKCTEGR	818		

```

OY      819 EFGSSRLVNLALPGIRFDKSDCODATYNLTGTVDLVRNSPPCCTTLRTISGDGMKTF   878
          |  ||| :||::||: ||| ||| :||: ||| ||| :||: ||| ||| :||: |||
Db       819 SFDDSNLFNLEPLPGIVKREKESDCNDSEYDILTSLVPILRDPCLTRINDPKCTTALVSGASWETI   878
QY      879 GTNLARQALVLDRAGNHFCFNSNFEEAFSOFSEFLRGS   914
          ||| ||| :||: ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       879 ANNLAHQALQVRAGSHVAFSPMFEVLGGQVEFEVRGS   914

RESULT         4
092398         PRELIMINARY:           PRT:    928 AA.

AC      092398
DT      01-MAY-1999 (TREMBLrel. 10, Created)
DT      01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT      01-MAY-2000 (TREMBLrel. 13, last annotation update)
DE      OUTER MEMBRANE PROTEIN PRECURSOR.
GN      OMP10 OR PMP_9.
OS      Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC      Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-CWL029.
RX      MEDLINE; 99206606.
RA      Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RT      Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RL      "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RN      Nat. genet. 21:385-389(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-VRI310;
RX      Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA      Christensen G., Birkebaek S.;
RT      "Chlamydia proteins containing the GGA1-repeat belong to a subfamily
RL      of autoantigenic pathogenicity factors.";
RN      Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-J18;
RX      Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
RA      Takeuchi A., Nishida J., Shibata K., Fujinaga R., Yoneda H.,
RA      Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,
RA      Ishii K., Shiba T., Hattori M., Kuhara S.;
RT      "Comparison of outer membrane protein genes omp and pmp in the whole
RT      genome sequences of Chlamydia pneumoniae isolates from Japan and
RT      US.";
RL      Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AE001628; AAD1891.1; -.
DR      EMBL; AJ133034; CAB37069.1; -.
DR      EMBL; AB033803; BAA85956.1; -.
KW      Signal.
FT      SIGNAL                     1               27              POTENTIAL.
FT      CHAIN                      28             928              OUTER MEMBRANE PROTEIN.
SO      SEQUENCE     928 AA;  98332 MW;  56910A8F04F12219 CRC64;

Query Match                41.5%; Score 1982; DB 2; Length 928;
Best Local Similarity      45.2%; Pred. No. 1.3e-103;
Matches 429; Conservative 150; Mismatches 327; Indels 44; Gaps 19;

QY      1 MKSSFPRKFVFSTFAIEPLSM-----IATEIVLDSSASFQDKNGCNFSVRESQEDA-GTT   53
          ||| | :|: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db       1 MKSSLHMFLLISSIALPLSLNFSFAFAAVEINELGTNTSFSG--PETYTPPAOTRWADGRI   58
QY      54 YLFKNVTLENIPGCGTAITKSCFNNTGDLITFGNGNSLFPQTVDAGTVACAANSSVV   113
          |  | :|: | :| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       59 YNLTDVDSITN-AGSPGTLTALASCPEETGGNLSFOGHGYQLLQNIDAG--ACCTFTNTPA   115
OY      114 DKSTFIFFSSLSPIASGSSITTKRGAVCSGTSLSTLNKVSLSFKNSTPDNGCAITA   173
          :| :||| ||| :| :||| ||| :||| ||| :||| ||| :||| ||| :||| |||
Db       116 NKLSLFSFSYSLIQT--TNATTGTGAIK-STGACGSIQSNNSCIFGQNFSDNGCALOG   172

```

```
QY 174 KTLISLTGTTMSALFSENTSSKKGGAIoTSDALHTTGNQGEVSPSDNTSSDGAIFTEAS 233
DB 173 SSSISLS-LMPNLTFAKNKATOKGALYSTGGITINNLTMSASEFENATPAANGCAITTEAS 231
QY 234 VTISNNKVSFIDNKVTGASSTTGMGGAICAYKTSIDTKV-LTLTGNOMLFFSNNTST 292
DB 232 SFTSSNKRAISFTINNSVTATSA-----TGCAIYCSSTSAKPPVLLTSDNGELNFIINTAI 285
QY 293 TGGGAIYVKKLELASGGLLFFSRNSVNGTARKGAIALIEDSELSSADSGDIVLQNT 352
DB 286 TGGGAIYVKNLVLSSGGLPLFNKNSAIDTAPALAIADSGSLSLSLGGLDITFEQNT 345
QY 353 V---TSTPGTTRSSIDL-TSAKMTALRSAGRAIYFDPITTGSTVTVDLKVNEPT 408
DB 346 VYKGAASSGOTTNRNSININTNAKIYQLASOGNTIYFDPIITSTITALSDALNNGPD 405
QY 409 ADSALOYTNIIFTGKELSETPEADSKNLTSKLQPVLTSGGTLSSKHGVTLOTQAFTQ 468
DB 406 LAGNPAYOCTIYFSGEKLEAEAEADNLKSTLOOPTLAGGOLSLKSGVTLVAKSFSS 465
QY 469 ADSRLEMVGTLEPADSTINNVLINISSIDGAKAKIETAKSKNLTLSGTLITLDP 528
DB 466 POSTLMDGTTLETADGTTITNNVLNVDLSKETKATKATQASOTVTLSSLSLVDS 525
QY 529 GTFYEHSILRNPOSYDILELKAS--GTVSTAVTPDPIMGEKFEHYGQGTWGPVWGTA 586
DB 526 GNVYEDVSNNNQVFSCLLTADDPANIHITDLADPLKKNPIHMYQGNMA-LSMQEOT 584
QY 587 ST---TATNMTKTGYTPNPERIGSLVPSLWNAFLDISLHYMETANEGLQGDRAFC 643
DB 585 ATRSKAATLTWKTGYNPBERGTLVANTLWGSFVDVRSIOOLVATKROSQETRGINC 644
QY 644 AALSNFPHDSTKTRGFEHLSGGVYIGGNLHMCSPKIIISAFCQLEFGDRDYFAKANG 703
DB 645 EISINFEHKOSTRKINKGFHISAGIYVGAITTLASDNLITLAAFCQLEFGDRHFIKNKA 704
QY 704 TVYGGTLVYOHNETYISLPCKLRPCSLSYV---TEIPVLFSGNLSYTTDMLTKYTT 760
DB 705 SAYAASLHLOHIALTSS-----PSLRLYLPQSESEQPLVPAQIISYISKMTMYIQ 758
QY 761 YTVVGSWGNDSFALFEGGRAP-ICIDESALFEQYMPFKLOFVYAHQGEFQGT-AR 818
DB 759 ARKGSWYNDCACALASLSPHALSHEDLHAFFPFIKEVASYIHQSFERNITLVR 818
QY 819 EFGSRLVNLALPIGIRPKESDCODATYNLTGYTVDLVRSNPDOCTTLRLSGDSKTF 878
DB 819 SFDSDGLIVASVPIGITFERFSRNERASTYEAIVYADVYRKPNDOCTTALLINNTSWKT 878
QY 879 GTNLARQALVLRAGNHFCFNSNFEAFSOFSEFLRGSSRNYNVDLGAQYOF 928
DB 879 GTNLSRQAGIGRAGIFYAFSPMLEVTSNLSMEIRGSSRSYNDLGGKQCF 928

RESULT 5
QY2898 PRELIMINARY: PRT: 936 AA.
AC Q92898:
DT 01-MAY-1999 (Tremblrel. 10. Created)
DT 01-MAY-1999 (Tremblrel. 10. Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13. Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
GN PMP-7.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029.
RX MEDLINE: 99206606.
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.C., Trachomatis.",
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.",
RL Nat. Genet. 21:385-389(1999).
DR EMBL; AF001627; AAD18589.1; -.
```

```
SEQ SEQUENCE 936 AA: 100079 MW: 8881D78A53D194BC CRC64;
Query Match 41.2%; Score 1965; DB 2; Length 936;
Best Local Similarity 43.4%; Pred. No. 1,2e-102;
Matches 411; Conservative 179; Mismatches 327; Indels 30; Gaps 16;

QY 1 MKSSFPEKVFSTFAIF-FLSMATETVLDSS-ASFDGKKNQNFVSRESQF-DAGTTLFFK 57
DB 1 MKSSVSWLFFSSIPLFSSLSIVAAEVLTDSSNNSYDGSNGTTFVFTSTDAAGTYSLL 60
QY 58 GNVLTENIPGTATKTCFNTKGTDLFTFGNGSLLTQVADAGVAGAAVNSVVDKST 117
DB 61 SVYSPQNNAGALGPIPLASCFLEAGDGLTFQGNQNAKRAFAINAGSAGTAVSTAADKNL 120
QY 118 TFIGSSLSFTASPGSSIT-7GKAVSCSTGSLSTFKVNSLFSKNESTDNGAITYAKTL 176
DB 121 LFNDFSRLSIISCPSELSTPTGOCALK-SVGNLSLTGNSQIIFQNFSSDNGVINTKNE 179
QY 177 SLTGTMSALFSENT--SSKKGGAIoTSDALHTTGNQGEVSPSDNTSSDGAIFTEASV 234
DB 180 LLSGTISQFASFSRNOAFTGKGGVYATGTTIENSPEIVSFQWLALAGSGALSTIDNC 239
QY 235 TISNNKVSFIDNKVTGASSTTGMGGAICAYKTSIDTKVTLTGNOMLFFSNNTSTTA 294
DB 240 SITDNFQYIFPGNSAMEAAQA-----QGGAICG--TTTDKVTTLGNKNLSFTNNLTALY 292
QY 295 GGAITYVKKLELASGGLLFFSRNSVNGTAPK--GGAIALIEDSGELSSADSGDIVLQNT 352
DB 293 GGAISGLKLVSIAGGPTLFOSN-LSGSSAGCGGGGAIINIAAGALASATSGDITFNNNQ 351
QY 353 VTSTPGTNRSSIDLGTSAKMTALRSAGRAIYFDPITTGSTTVDVLVNEFPADSA 412
DB 352 VTNGSTST-RNAINIIDTAKVTSIAATGOSIIFYDPIITNPGTAASTDTLNLNADASE 410
QY 413 LQYTNIIFTGKELSETPEADSKNLTSLKLOPVLSSGTLSSKHGVTLOTQAFTQADSR 472
DB 411 IEYGAIVFSGEKSLPTKAIANVTSTIRQPAVLARGDLVLRGVYTFFDLQSPSR 470
QY 473 LEMDVGTTLEPADTS-TNNVLINISSIDGAKKAKIETAKSKNLTLSGTLTLDPTGTF 531
DB 471 ILMDGTTLSAKKANLSLNGLAVNLSLDGINKAALKTEADKNISLSGTALDITDESF 530
QY 532 YENHSLRNPOSYDILELK---ASGVTSTAVTPDPIMGEKFEHYGQGTWGPVWGTA 588
DB 531 YENHSLKASSTYPILELTAGANGITIGALSTLTLDQEPHYGQGNW-QLSMANATSS 589
QY 589 -TAFNMTKTGYTPNPERIGSLVPSLWNAFLDISLHYMETANEGLQGDRAFCAGLS 647
DB 590 KIGSINMTRTGYISPERKSNLPLNSLWGNFIDIRSIINOLIETKSSGPFERELWLSGIA 649
QY 648 NFEHKDSTKTRGPRHLSGIVIGGNLHTCSDKILSAFCQLFEGDRDYFAKNOGIYV 707
DB 650 NFEYRDSMPTRRGFRHISGVALGTTATTPAEDQLTFAFCQLFARDRHITGKNHGDY 709
QY 708 GLIYVQHNETYISLPCKL-----RPCSLSYVPTIEIPVLFSGNLSYTTDMLTKYTTY 761
DB 710 ASLIVHHNEGFLDIANFLMGKATRAPWLSIISQIITPLSPAKRSYLTHTDHMTYYTIDN 769
QY 762 PTYKGSWGNDSFALFEGGRADICIDESALFEQYMPFKLOFVYAHQGEFQGTIAREFG 821
DB 770 SIKGSWRNDAFCADLGSALPFIYSVPLLKEVEPFAVQYIVYAHQGEFVRAHAGRAFN 829
QY 822 SSRVLNLAIPGIRPKESDCODATYNLTGYTVDLVRSNPDOCTTLRLSGDSKKTGTN 881
DB 830 KSELINVEIPIGVTFERDSKSEKGTIDYLLMYIIDAARRNKCOTSLASDANMAAYGTN 889
QY 882 LARQALVLRAGNHFCFNSNFEAFSOFSEFLRGSSRNYNVDLGAQYOF 928
DB 890 LARQGSVRRANHFQVNPHEITFGQAFEVWSSSRNNTNLGSKFCF 936

RESULT 6
```

09R866
ID 09R866 PRELIMINARY: PRT: 930 AA.
AC 09R866:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DE POLYORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
GN PMP_8.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138.
RA Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
RA Takeuchi A., Nishida J., Shibata K., Fujinaga R., Yoneda H.,
RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,
RA Ishii K., Shiba T., Hattori M., Kuhara S.:
RT "Comparison of outer membrane protein genes omp and pmp in the whole
genome sequences of Chlamydia pneumoniae isolates from Japan and US";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB033802; BAA85955.1; -
RL SEQUENCE 930 AA: 97639 MW: 46a8896761391C09 CRC64;

Query Match 40.6%; Score 1940; DB 2; Length 930;
Best Local Similarity 43.7%; Pred. No. 2.9e-101;
Matches 416; Conservative 165; Mismatches 324; Indels 48; Gaps 17;

QY 1 MKSSPFKFEVSTFAIPLSM-IAT---ETVLDSASFDGKNGNFVSRESQEDAGTYLIF 56
DB 1 MKPLKLLISSFLVPIILSTIATYGADASLPTDSFDGAGSGSTFTPK\$ADANGTYVL 60
QY 57 KGNVLTLENIPGDTAATKSCFNNTKGDLTFTGNGNSLFTQVDAGVAGAAVSSVVDKS 116
DB 61 SGNVYI-NDAGKGTALGCCFTETGDLFTGKGYSEFNTVADGSAVAGAA-STRADKA 118
QY 117 TTFIGFSSLFASPGSSITFGKAVSCSTGSLTKRVNLSLFSKNSPD---NGGATTA 123
DB 119 LFTGFSNLSFLIAPGTIVASGKSTLS-SAGALNLNDNGTILTSQNVSEANNNGATTA 177
QY 174 KTLISLTGTTMSALFSENTSSKKGAIQIOTSDALTITGNOGEVSEFSDNTSSPGAAITFEAS 233
DB 178 KTLISGNTSITFTSNAKKGAIYSSAASISGVTGQLVNNKKGTEGGALGEAS 237
QY 234 VTISSNAKVSFIDNKTGASSTTGDMSCGALCAIKTSTKTLTGNGQMLLSNNTST 293
DB 238 SSTITQNSLFEFSGNTATDAG-----KGAIIICEKGETPTLTISGNKSLTEAENS SVT 291
QY 294 AGGAIYVKKLELASGGLTFSRNSVNGTTPKGAIAIEDSGELSLADSGDIVFGNTV 353
DB 292 QGGAICAHGLDLSAAGPTLFSNNRCGNTAAGKGAIALADSGLSLSAAGDITFLCNTL 351
QY 354 TSTT-PTNNSSIDLGTSAKMTALRSAGRAIFYDPI---TTGSSITTVIVLVKVEIPA 409
DB 352 TSTTAPTSTRNAIYLSSAKITMLRAAGOSIFYEYDPIASNTGAS---DVLINOPDS 407
QY 410 DSALOTGNTIIFGKELSTFEADSKNLTKLQPYTLTSGTISLKHGTYLQTAFTQQA 469
DB 408 NSPLDYSGLTIVFSGEKLSADEAKAADFTSLKOPALAGLALGKVELDNGVFOYE 467
QY 470 DSRLEMDVGTLEPADTSTIN-NLVNISSIGAKKAKETKATSKNLSTGTLTLP 527
DB 468 GSTLMDQPTKRL-ADTEALSTLKLIVDLSALGKNKSVSETAGANKTITLISPLVPODS 526
QY 528 TGTFFYNHSLRNPOSYDIELKAS-----GYTSTAIVPDPPIGKEKFEHYGQGTWG 578
DB 527 SCNFYSHTINQAFQPLVYFTATAASDIYIDALLITSPTQTEP-----HYGYGHW 580
QY 579 PIWGGASTATFENMTKTGYIPNPERIGSLVNSLMNADISSLHYMETANEGLQGD 638
DB 581 ATAAADSTASGIMVTWYTYINPERRASVVDLSMASTDIRILOITQANSIYOO 640
QY 639 RAFWAGLANSFHKDSTKTRGRGFRHLSGGYVIGNLTGCSDKILSAAFCOLFGRDRQYFV 698

DB 641 RGLMAGSTANFHKDKSGTNOAFRRHKSXYIVGGSNADESENIFFSAFQGLGKDKDLFI 700
QY 699 AKNOGVYCGTLYYONENETIYSLPCKLRPC--SLSVPTPEIVLFGNSLTYHTNDLKT 756
DB 701 VENTSINYLASLYLQHRALGLGP---MPSFSTIDMLKDIFLIINAQISYTKNDMT 757
QY 757 KYTTPYVKSGMDSFALEFGRAFICL-DESALEQYMPFMKLOFYVAHOGEFKEQGT 815
DB 758 RYTSYEPAGOSWTNNGALGSLALYLPKPAFFQGYFPLKQAVYSROONFRESGA 817
QY 816 EAREFSSRLVNLALPIGIRPDKESDCQATYNTLIGYTVDLVRSNPDCTTLIRISGDSW 875
DB 818 EARAEPDGLVNCISIPVGIRLEKISEDEKNNEISLATYIGDYRRKPRSRISLWSGASW 877
QY 876 KTFGTMLARQALVLRAGNHFECNSNFARSQSFELRGSSRYNVNDIGAKYOF 928
DB 878 TSLCKLAAQAFLASGSHLITLSPHELSEAAVELRSGAHYINVDGGLRYSF 930

RESULT 7
092393
ID 092393 PRELIMINARY: PRT: 930 AA.

AC 092393:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DE OUTER MEMBRANE PROTEIN 11 PRECURSOR.
GN Omp11 OR PMP 8.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029.
RX MEDLINE: 99206606.
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.:
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
RT Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310.
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkelund S.:
RT "Chlamydia proteins containing the GCAT-repeat belong to a subfamily
of autotransporting pathogenicity factors";
RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE001627; AAD18590.1; -
DR EMBL: AJ133034; CAB37068.1; -
FT Signal.
FT SIGNAL 1 27
FT CHAIN 28 930
FT SIGNAL 28 930
SQ SEQUENCE 930 AA: 97669 MW: 46a9B5E3BB913C4C CRC64;

Query Match 40.6%; Score 1936; DB 2; Length 930;
Best Local Similarity 43.5%; Pred. No. 4.9e-101;
Matches 415; Conservative 165; Mismatches 325; Indels 48; Gaps 17;

QY 1 MKSSPFKFEVSTFAIPLSM-IAT---ETVLDSASFDGKNGNFVSRESQEDAGTYLIF 56
DB 1 MKPLKLLISSFLVPIILSTIATYGADASLPTDSFDGAGSGSTFTPK\$ADANGTYVL 60
QY 57 KGNVLTLENIPGDTAATKSCFNNTKGDLTFTGNGNSLFTQVDAGVAGAAVSSVVDKS 116
DB 61 SGNVYI-NDAGKGTALGCCFTETGDLFTGKGYSEFNTVADGSAVAGAA-STRADKA 118
QY 117 TTFIGFSSLFASPGSSITFGKAVSCSTGSLTKRVNLSLFSKNSPD---NGGATTA 123
DB 119 LFTGFSNLSFLIAPGTIVASGKSTLS-SAGALNLNDNGTILTSQNVSEANNNGATTA 177
QY 174 KTLISLTGTTMSALFSENTSSKKGAIQIOTSDALTITGNOGEVSEFSDNTSSPGAAITFEAS 233
DB 178 KTLISGNTSITFTSNAKKGAIYSSAASISGVTGQLVNNKKGTEGGALGEAS 237
QY 234 VTISSNAKVSFIDNKTGASSTTGDMSCGALCAIKTSTKTLTGNGQMLLSNNTST 293
DB 238 SSTITQNSLFEFSGNTATDAG-----KGAIIICEKGETPTLTISGNKSLTEAENS SVT 291
QY 294 AGGAIYVKKLELASGGLTFSRNSVNGTTPKGAIAIEDSGELSLADSGDIVFGNTV 353
DB 292 QGGAICAHGLDLSAAGPTLFSNNRCGNTAAGKGAIALADSGLSLSAAGDITFLCNTL 351
QY 354 TSTT-PTNNSSIDLGTSAKMTALRSAGRAIFYDPI---TTGSSITTVIVLVKVEIPA 409
DB 352 TSTTAPTSTRNAIYLSSAKITMLRAAGOSIFYEYDPIASNTGAS---DVLINOPDS 407
QY 410 DSALOTGNTIIFGKELSTFEADSKNLTKLQPYTLTSGTISLKHGTYLQTAFTQQA 469
DB 408 NSPLDYSGLTIVFSGEKLSADEAKAADFTSLKOPALAGLALGKVELDNGVFOYE 467
QY 470 DSRLEMDVGTLEPADTSTIN-NLVNISSIGAKKAKETKATSKNLSTGTLTLP 527
DB 468 GSTLMDQPTKRL-ADTEALSTLKLIVDLSALGKNKSVSETAGANKTITLISPLVPODS 526
QY 528 TGTFFYNHSLRNPOSYDIELKAS-----GYTSTAIVPDPPIGKEKFEHYGQGTWG 578
DB 527 SCNFYSHTINQAFQPLVYFTATAASDIYIDALLITSPTQTEP-----HYGYGHW 580
QY 579 PIWGGASTATFENMTKTGYIPNPERIGSLVNSLMNADISSLHYMETANEGLQGD 638
DB 581 ATAAADSTASGIMVTWYTYINPERRASVVDLSMASTDIRILOITQANSIYOO 640
QY 639 RAFWAGLANSFHKDSTKTRGRGFRHLSGGYVIGNLTGCSDKILSAAFCOLFGRDRQYFV 698

```

Db 178 KTLISGNTSSITFTSNAGAKKLGAIYSSAAASISGNTGQLVPMNNKGETGGALGFEAS 237
QY 234 VTISNNAKVSFIDNKVGTGASSSTTGDMGGAICAKITDTRKVLITGOMLLFSNITST 293
Db 238 SSITONSSLEFSGNATPAAG-----KGAIYCEKTEIPLTITISGNKSLTFEENSSTV 291
QY 294 AGGAIYVKKLELASGTLTFSRNSVNGTAPKGAIAIEDSGELSLADSGIVLQMTV 353
Db 292 OGGAICAHGLDLSAAGPLTFSNNRCGNTAAGGAIAIADSGSLSLANOGDITLGLMTL 351
QY 354 TSTT-PGTRSSIDLGTSAKMTALRSAGRAIYFYDP1---TTGSSITVTDVLKVNTPA 409
Db 352 TSTSAPTSRNMIYIGSSAKITNLRAAQGSITFYDPIASNTTAS----DVLITNGDS 407
QY 410 DSALOYGNITFTGEKLESTELPADSKNLTKLLOPYTLSGGLTSLKHGYLTQTAFTQQA 469
Db 408 NSPLDYSGTIVFSGEKLSDAEKADNFTSILKOPALASGLLAKGNELVDNGFTOTE 467
QY 470 DSRLEMDVGTLEPADSTIN--NLVINISSIDGAKKAKIEFKATSKNLTLSGTTLLDP 527
Db 468 GSTLMQPTKAK-ADTEAISTKLIVDLSALEGNKSVSIEFAGANKTITLTSPLVFQDS 526
QY 528 TGTFEYENHSLRNPQSYDILELKAS-----GVTSTAVTPDPIMGEKHYGTQW 578
Db 527 SGNFYESHTINOAFQTPLVVFTAATPAADYIDALLTSFVOTPEP-----HYGYQGHME 580
QY 579 PLVWGGAATTAFTNNTKGYIPNPERISLYPNLSIMNAFIDISSIHYLMETANGLQD 638
Db 581 ATWADISTAKSGMTVTTGYNPNPERASVVPDSLMASTDIRILOQITWISOANSITQ 640
QY 639 RAFCWGLSNFHKDSTKTRRGFRHLSGYVIGNLTMSDKILSAFCOLFRDRDYFV 698
Db 641 RGLMAGSTANFHKDKSGINQAFRHSYGYIVGSADEPSENFVAFCQLGKXDLFI 700
QY 699 AKNOGTVGGTLYYQHNETYISLPCKLRPC--SLSYVPEIPIVLSGNLSYTHINDLKT 756
Db 701 VENTSHNYLASLYLQRRARELGLP--MPSFGSITDMLDPLILNAQSYSTYKNDMT 757
QY 757 KYTTYTVGSGMNDFALEFGRAPICL-DEGALFEQYMPMKLOFYVAHDEGFEQGT 815
Db 758 RYTSYPEAGSGWNNSGALEGSLALYLPKEAPFQGYFPFLKFOAYVSRQONKESGA 817
QY 816 EAREFSSRLVNLALPIGIRFDEKSCODATYVNLGYTVDLVRSPDCOTTLRISGDM 875
Db 818 EAAFPDGLVNCISPIVGRLEKISDEKNNPEISLAYIGDYRKRPRTSLMVGASW 877
QY 876 KYTGTNLARQALVLRAGNHFCEFSNFEAFSOFSEFELRGSSRYNVDLAKAYOF 928
Db 878 TSICKNLARQAFLASAGSHLLTSPHYELSGEAYELRGSNHIYVDCGRYSP 930

RESULT 8
P71135 PRELIMINARY; PRT; 926 AA.
AC P71135;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)
OS PUTATIVE 98 KDA OUTER MEMBRANE PROTEIN.
OC Chlamydia phila psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ABORTION S26/3;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U72499; AAB18188.1;
SQ SEQUENCE 926 AA: 98439 MW: 3E75E52F594750F CRC64;

```

Query Match 40.18; Score 1915; DB 2; Length 926;
 Best Local Similarity 42.9%; Pred. No. 7.4e-100;
 Matches 410; Conservative 156; Mismatches 333; Indels 56; Gaps 18;

```

QY 1 MKSFPKVESTFAIFPLSW-----I 1
Db 1 MRSPLKILLISSTLIPISFHFSQLHAVAL 1
QY 51 GTTYLFKNVNTLENIPOTGTAITKSCFNMTK 51
Db 53 GTIYNVEDSIDIYDV-QOTAAALSAFVOJA 53
QY 111 SYVDKSTTFIGFSSLSFTIASPSSITTKGZ 111
Db 111 NTAKILTLFDPSKLSKECPSSLVNTKGZ 111
QY 171 ITAKTLSTGTTMSALPESNTSSKKGAIQ 171
Db 170 ISCAFAFLTSSKEISFTTNTAKKGAIK 170
QY 231 EASYTISNNAKVSFIDNKVGTGASSSTTGD 231
Db 230 EASMTIAGNNHVAFSNNAVSGSS-----DCCGGAHCSKIGSARF 230
QY 291 STTAGAIYVKKLELASGGLTLFSRNSVNGTAPKGAIAIEDSGELSLADSGDIVFLG 350
Db 285 SSAGGAIYIDKLITLSGPTAFINKVTHAT--PKGAICIAANGECSLAEHGDITFON 343
QY 351 NTVSTTPTG-TRSSIDLGTSAMTALRSAGRAIYFYDPIITGSSITVTDVLKVNTPA 409
Db 344 NLMTQDNATIKRNAVINEGKGFVNLRASGKTISEYDIIVEGN--AADLLTLNKAEG 401
QY 410 DSALOYGNITFTGEKLESTELPADSKNLTKLLOPYTLSGGLTSLKHGYLTQTAFTQQA 469
Db 402 DKT--YNGRIIFESGKLEEOAAVADNLKTTFTOPTIILAGELVLSGVEKAYVQTA 459
QY 470 DSRLEMDVGTLE-PADSTINNVLINISSIDGAKKAKIEFKATSKNLTLSGTTLLDP 528
Db 460 GSLIMDAGTSLSKATEDALITNLAINPNTLDGKRFVDAVAAGKAVTLSGALGVIDPT 519
QY 529 GTFEYENHSLRNPQSYDILELKASGTVTSTAVTPDPIMG-EKFIHYGTQWGPVW----- 582
Db 520 GKFEYENHKLNDTLAGIQLSGKGSVTTINV-PSHVGVAEETHGYOGNWS-VSWKDN 577
QY 583 GTGASTATFTNWTGTGYTPNPERIGSLVPSIMNAFIDISSIHYLMETANGLQD 641
Db 578 SDPKQTAIFTNKNTGYVNPERRAPLVNLSMGSTFDLNSIODVLEKSDLSLETRGL 637
QY 642 WCAGLSNFFHKDSTKTRRGFRHLSGYVIGNLTMSDKILSAFCOLFRDRDYFAKN 701
Db 638 WVSIGNFFHKDRNAENKFRHISGSYVLGATINTSHEDLSVAFQCLPAKDKDYLSKN 697
QY 702 QGTYYGGLTYQHNETYISL-----PCKLRPCSLSYVTEIPIVLSGNLSYTHIND 753
Db 698 AANYAGGSVYQHVSKFPDDLTLRFNGPNTC-----CSGFSKEIPIPLDAQIITYCHTANN 751
QY 754 LKTKYTPYKSGMNDFALEFGRAPICLDESALFEQYMPMKLOFYVAHDEGFEQ 813
Db 752 MTTSTYDTPVKGSGMNDTGLTLTSTVPIPVFSSSIFDSTAPAKQOYVYAHODDEK 811
QY 814 GTEAREFSSRLVNLALPIGIRFDEKSCODATYVNLGYTVDLVRSPDCOTTLRISGD 873
Db 812 TTEGVFESSDLNVSIVIGIKFEKLSYGERSAVDLTLMYIPDYRRNPMCGLAINDV 871
QY 874 SWKTFGNLARQALVLRAGNHFCEFSNFEAFSOFSEFELRGSSRYNVDLAKAYOF 928
Db 872 SWLTATNLARQAFIVRAGNHIALTSGVEMFSQGFELRSSRYNVDLAKAYAF 926

RESULT 9
O86164 PRELIMINARY; PRT; 928 AA.
AC O86164;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)

```


DE OUTER MEMBRANE PROTEIN 4 PRECURSOR.
GN OMP4 OR PMP-11.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC/CW-029/VR-1310.
RA Knudsen K., Madsen A.S., Myind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding outer membrane complex
associated surface layer proteins in Chlamydia pneumoniae."
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VR1310.
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RT Christiansen G., Birkelund S.;
RT "Chlamydia proteins containing the GCAI-repeat belong to a subfamily
of autotransporting pathogenicity factors."
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CW1029;
RC MEDLINE: 99206606.
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Gilwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RT Nat. Genet. 21:385-389(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RA Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
RA Takuchi A., Nishida J., Shibata K., Fujinaga R., Yoneda H.,
RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,
RA Ishii K., Shida T., Hattori M., Kunara S.;
RT "Comparison of outer membrane protein genes omp and pmp in the whole
RT genome sequences of Chlamydia pneumoniae isolates from Japan and
US."
RT Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [5]
RP EMBL: AJ001311; CA004672.1; -;
DR EMBL: AJ133034; CAB37072.1; -;
DR EMBL: AE001628; AAD18593.1; -;
DR EMBL: AB033806; BAA85959.1; -;
KW Signal.
FT SIGNAL.
FT CHAIN 1 17 POTENTIAL.
FT SEQUENCE 928 AA, 98903 MW, 7888BCDD62C911402 CRC64;

Query Match 38.9%; Score 1855; DB 2; Length 928;
Best Local Similarity 42.7%; Pred. No. 1.7e-96;
Matches 405; Conservative 171; Mismatches 331; Indels 42; Gaps 20;

QY 1 MSSPFKVFSTFAIRP--LSMATITVYDSSASPDGN-KNGFVSREQEDAGTYLTK 57
DB 1 MKTSLPWFVSVSLFASCHLQSLANELLSPPDSFNGNIDSGFTFKTS---ATVYSLT 56
OY 58 GAVTLENIGTGTAIRKSGFNNTKGLDFTFGNGNSLTFQTVAGATAGAAVSVVDKST 117
DB 57 GGVFFYE-PKGTPLSDSCFKOTTDNLFGLNGHSLTFGFDIDGTHAGAAA-STANKKL 114
OY 118 TFGFSSLSFLIAPGSSITFGKAVSCSTGSLTKNLSLFKNSFTNDNGAIFAKTIS 177
DB 115 TFGFSSLSFLIAPGSSITFGKAVSCSTGSLTKNLSLFKNSFTNDNGAIFAKTIS 173
OY 178 LCTGTSALEFSEKTSKKGATOTSDALITITGNGEVSNDNTSSSGAIFTEASVTIS 237
DB 174 LCTGTSALEFSEKTSKKGATOTSDALITITGNGEVSNDNTSSSGAIFTEASVTIS 233
OY 238 NNAKVFINKVYTDGSSSTGDMGCAICAYKTSTDTKTYLGNQMLLSNNSTTAGGA 297
DB 234 NKKFLVF-----EGNAKTT-----GGAICNTKASSGPELLISNNKLLIRASNAVETSGGA 284
OY 298 IYVKLELASGGLTLFSRNSVNGGTAPKGAIAIEDSGELSLSDSGDIVFLGNTVTSF- 356

DB 285 IIAKKLASGGFTEFLRNVSSAT-PKGAISIDASELSIATETGNTITVRNLTITG 343
OY 357 -TPGNRSSIDGTSAKMTALRSAGRAIYFDPITTSSTVTIVLVNENPADAQY 415
DB 344 STDTPKRNAINNGSKFTLPAKNNHITFFDPTT--SEGSSVLKINNRSALAPY 401
OY 416 TGNITFGKLETEADSKNLTSLKLOPVTLSGGTSLSKHGVTLOTQAFTOQADSRLM 475
DB 402 OCTIFSGCTFLADKLKADNLSSTFQPSVLSGGKLLQKQVLTSTFSQEAASLLGM 461
OY 476 DVGTFLE-PADSTNNVINISSIDGAKKAKIERKATSKNLTSGTITLLOPTEFEN 534
DB 462 DSGTILSTTAGSITITTNAGVNDISLGPVSLAKGASNNKYVSGKLNLDIEGNITES 521
OY 535 HSLRNPQSYDILELKASGVTS---TAVTPDPPIGEKFFHYGQGTWGPVWGASTT- 589
DB 522 HMFSDQLFSLKTIIVDADVQNDVNDISSILPVPADPSEYQFOGQWN-VNMTTITATN 580
OY 590 --ATFNWTKGTIPNPERIGSLVPSLWNAFLDISLSLHYLETANEGIQGDRAPACAGS 647
DB 581 KEATATWTKGTVPSPERSALVCNTLWGVFDIRSLQOLVEIGATGHEHKGCFVSSMT 640
OY 648 NFEHDKSTRGRGFHLSGGVYIGNLHPCSDKILSAFCOLFGDRDPYFAKNOGTYYG 707
DB 641 NPLHTGDENRKGFHRTSGGYIGSAHTPKDDLTFEAFCHLFAADKCFIAHNNSTYTG 700
OY 708 GTLYVQHNET-----YISL-PCKLRCSLSYVTEIPVLFSGNLSYTHDNDLTKTYTY 761
DB 701 GTLPFKSHHTLOPQVYLRIGRAFSESALIEKPREIPALDVOVSFSHSDNMETHYSL 760
OY 762 PIVKSWGDSFALFEGGRAPICL-DESALEQVWPEPKLOFVYAHOGCFEGEAEAF 820
DB 761 PSESSWSMECIAGIGIDLPVLISNPHLRTFIPQKVEVWYVYSONSFESSDGGCF 820
OY 821 GSSRLVNALPIGIFEDKESQDA-TYNLTIGYVTDVRSNPDCTTLRISGDSMKTFG 879
DB 821 STGRLLNLSIPVGAFF-VQGDIGDSYTYDLGSEFVSQVYRNNPOSTATLVSPDSKIRG 879
OY 880 TNLARQALVLRAGNFCFNSNFARFQSFELRGSSRNINVDLAKKYP 928
DB 880 GNLSRQAFELRGSSNNYVNSNCELFGHAMELRGSSRNINVDVGTKRF 928

RESULT 10
P77792 PRELIMINARY; PRT; 839 AA.
AC P77792;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE POMP908 PRECURSOR.
OS Chlamydia psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
RN [3]
RP EMBL: U65943; AAC15924.1; -;
DR EMBL: U65942; AAC15922.1; -;
KW Signal.
FT SIGNAL.
FT CHAIN 1 16 POTENTIAL.
FT SEQUENCE 839 AA, 89824 MW, 4581C7CBAF7FF4C4 CRC64;

Query Match 34.7%; Score 1656.5; DB 2; Length 839;

OM of: US-09-428-122-2 to: GenEmbl:* out_format : pfs

Date: Nov 20, 2000 4:36 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=itame+g2n.model -DEV=xlp  
-o=/cn2_1/USPTO_spool/US09428122/runat_14112000_163316_27911/app-query.fasta_1.999  
-B=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blissum62 -TRANS=human04.cdi  
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=2000000000 -USER=US09428122.@CGN1_1_5773 -NCPU=6  
-ICPU=3 -LONGLOG -NO_XLPPX -WAIT -THREADS=1
```

Search information block:

Query: US-09-428-122-2

Query length: 928

Database: GenEmbl:*

Database sequences: 1033670

Search time (sec): 2334.060000

Score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pat:AE002235	- 4760.00	4008.25	1.7e-216	2787	AB1835 Sequence 9 from Patent
gb_pat:AE002235	- 4746.00	4008.48	3.8e-215	12127	AE002235 Chlamydia pneumoniae
gb_pat:AE001587	+ 4746.00	4005.73	5.3e-215	16448	AE001587 Chlamydia pneumoniae
gb_pat:CPN133035	+ 4746.00	4001.60	9.0e-215	26920	AJ133035 Chlamydia pneumoniae
gb_pat:AP002545	+ 3865.00	3276.51	1.2e-174	300050	AP002545 Chlamydia pneumoniae
gb_pat:AB033795	+ 3865.00	3276.51	2.2e-174	2241	AB033795 Chlamydia pneumoniae
gb_pat:AB1829	+ 2058.00	1744.69	4.6e-89	2815	AB1829 Sequence 3 from Patent
gb_pat:AE002192	+ 2058.00	1732.09	2.3e-88	12676	AE002192 Chlamydia pneumoniae
gb_pat:CPN133034	- 2058.00	1729.49	3.3e-88	17280	AJ133034 Chlamydia pneumoniae
gb_pat:AP002546	- 2058.00	1705.58	7.0e-87	300600	AP002546 Chlamydia pneumoniae
gb_pat:AE001628	- 2044.00	1721.61	8.9e-88	10757	AE001628 Chlamydia pneumoniae
gb_pat:AB1837	+ 2001.00	1696.60	2.2e-86	2757	AB1837 Sequence 11 from Patent
gb_pat:AB1851	+ 2000.00	1695.05	2.7e-86	3000	AB1851 Sequence 25 from Patent
gb_pat:CPOMP4	- 2000.00	1689.20	5.7e-86	6030	AJ001311 Chlamydia pneumoniae
gb_pat:AB1839	+ 1986.00	1683.81	1.1e-85	2787	AB1839 Sequence 13 from Patent
gb_pat:AB033803	+ 1982.00	1680.43	1.8e-85	2787	AB033803 Chlamydia pneumoniae
gb_pat:AE001628	+ 1982.00	1669.12	7.5e-85	10757	AE001628 Chlamydia pneumoniae
gb_pat:AE002192	- 1982.00	1667.74	9.0e-85	12676	AE002192 Chlamydia pneumoniae
gb_pat:CPN133034	+ 1982.00	1665.15	1.2e-84	17280	AJ133034 Chlamydia pneumoniae
gb_pat:AP002546	+ 1982.00	1661.23	2.7e-84	300600	AP002546 Chlamydia pneumoniae
gb_pat:AE001627	+ 1965.00	1651.90	6.8e-84	15068	AE001627 Chlamydia pneumoniae
gb_pat:AB033849	+ 1964.00	1653.69	4.2e-84	8658	AB033849 Chlamydia pneumoniae
gb_pat:AE002193	- 1964.00	1654.47	4.9e-84	10026	AE002193 Chlamydia pneumoniae
gb_pat:AB033802	+ 1940.00	1644.85	1.7e-83	2793	AB033802 Chlamydia pneumoniae
gb_pat:CPU72499	+ 1927.00	1633.84	6.9e-83	2793	AB1841 Sequence 15 from Patent
gb_pat:AB1841	+ 1915.00	1618.93	4.7e-82	4926	U72499 Chlamydia pneumoniae
gb_pat:AB033806	+ 1885.00	1572.90	1.7e-79	2787	AB033806 Chlamydia pneumoniae
gb_pat:AB1827	+ 1855.00	1571.74	2.0e-79	3200	AB1827 Sequence 1 from Patent
gb_pat:CPOMP4	+ 1855.00	1566.44	3.9e-79	6030	AJ001311 Chlamydia pneumoniae
gb_pat:AB1849	+ 1811.00	1535.50	2.1e-77	2838	AB1849 Sequence 23 from Patent
gb_pat:CPU65942	+ 1659.50	1400.81	6.6e-70	6110	U65942 Chlamydia pneumoniae
gb_pat:CPU65943	+ 1659.50	1400.81	6.6e-70	6234	U65943 Chlamydia pneumoniae
gb_pat:AE001586	+ 1573.00	1322.98	1.4e-65	10574	AE001586 Chlamydia pneumoniae
gb_pat:AB1833	+ 1564.00	1323.35	8.2e-66	2526	AB1833 Sequence 7 from Patent
gb_pat:AB033793	+ 1560.00	1323.97	1.3e-66	2526	AB033793 Chlamydia pneumoniae
gb_pat:AB033792	+ 1442.50	1212.71	4.8e-60	2769	AB033792 Chlamydia pneumoniae
gb_pat:AE001585	+ 1442.50	1212.34	2.1e-59	10766	AE001585 Chlamydia pneumoniae
gb_pat:AE002237	- 1441.50	1212.07	2.2e-59	10044	AE002237 Chlamydia pneumoniae
gb_pat:AB1831	+ 1430.50	1212.73	2.0e-59	3052	AB1831 Sequence 5 from Patent
gb_pat:AB033808	+ 1377.50	1158.22	6.0e-57	2922	AB033808 Chlamydia pneumoniae
gb_pat:AE001629	+ 1377.50	1157.36	2.4e-56	10695	AE001629 Chlamydia pneumoniae

seq_name: gb_pat:AB1835

seq_documentation_block:

LOCUS	AB1835	2787 bp	DNA	PAT	21-JAN-2000
DEFINITION	Sequence 9 from Patent WO9858953.				
ACCESSION	AB1835				
VERSION	AB1835.1 GI:6731868				
KEYWORDS	unidentified.				
SOURCE	unclassified.				
ORGANISM	unclassified.				

REFERENCE
1 (bases 1 to 2787)
Madsen, A. and Birkelund, S.
NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
Patent: WO 9858953-A 30-DEC-1998;
MADSEN ANNA SØFIE (DK); BIRKELUND SVEND (DK)

FEATURES

source
1..2787
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 811 a 583 c 598 g 795 t

ORIGIN

alignment_scores:

Quality: 4760.00 Length: 928
Ratio: 5.135 Gaps: 0
Percent Similarity: 99.892 Percent Identity: 99.677

alignment_block:

US-09-428-122-2 x AB1835

Align seg 1/1 to: AB1835 from: 1 to: 2787

1	MetlyserSerPheProlySpheValPheSerThrPheAlaIlePhePr	17
1	ATGAAGCTCTTCTTCCCAAGTTGTAATTTCTACATTCTCTATTTCC	50
17	cleSerMetIleAlaThrGluThrValIleAspSerSerAlaSerPhe	34
51	TTTGTCTATGATTTCTACCGACAGTTTGATTCAGTCCAGTTTCG	100
34	spGlyAsnLysAsnGlyAsnPheSerValArgIuSerGlnIuAspAla	50
101	ATGGGAAATAAATAAGTAATTTTCACTTCGTGAGAGTCAGAAATGCT	150
51	GlyThrThrTrpLeuPheLysGlyAsnValThrLeuGlnAsnIleProG	67
151	GGAACTACCTCTATTAAGGAATGCTCTGAGAAATATATCTCTGG	200
67	YThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAsp	84
201	AACAGGACAGCAATCAACAAAAGCTCTTTAACACACATGAAGCGGAT	250
84	euthrPheThrGlyAsnGlyAsnSerLeuPheGlnThrValAspAla	100
251	TGACTTTCACAGGTAACGAGACCTCTATTGTTCCAAAGGTGATGCA	300
101	GlyThrValIleGlyAlaValIleAsnSerValValAspLysSerThr	117
301	GGAGCTTACAGGCGCTGCTGTTAACAGCAGCGTGTATATTAATCTAC	350
117	rThrPheIleGlyPheSerLeuSerPheIleAlaSerProGlySerS	134
351	CAGCTTATAGGGTTTCTTCGCTATCTTTATTCGCTCTCGAAGTT	400
134	erIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSer	150

401 CGATACTACCGCAAGAGCCGTTAGCTGCTGCTAGCGGTAGCTTGAG 450
 151 LeuThrIysAsnValSerLeuLeuPheSerIysAsnPhSerThrAspAs 167
 451 TTGACAAAATGTCAGTTGCTCTTCAGCAAAACCTTTTCACGGAAVA 500
 167 ngIyGIAlaIleThrAlaIylSerLeuSerLeuThrGlyThrMetS 184
 501 TGGGGGTGCTATCACCGCAAAACCTTTTCATTAACAGGACTCAATGT 550
 184 eAlaLeuPheSerGIuAsnThrSerIysGlyGlyAlaIleGln 200
 551 CAGCTGCTTTTCTGAAATACCTCTCAAGAAAGGGGAGCCATTGAG 600
 201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyIuValSerPhe 217
 601 ACTTCGAGTGCCTTACCATTTACTGGAACCAAGGAGGAGCTCTTTTC 650
 217 rAspAsnThrSerSerAspSerGIyAlaAlaIlePheThrGluIaSerV 234
 651 TGACAAATACTTCTTCGGATCTGGAGCTGCAATTTTACAGAACCTCGG 700
 234 aLThrIleSerAsnAsnAlaIylSerPheIleAspAsnIylValThr 250
 701 TGACTATTTCTAAATATGCTAAAGTTTCTTATTGACAAATAGCTACA 750
 251 GlyAlaSerSerSerThrThrGlyAspMetSerGIyGlyAlaIleCysAl 267
 751 GGACCGAGCTCTCTACACACGGGGATATGTCAGAGAGTGTATCTGTGC 800
 267 aTYrLYSThSerThrAspThrIylValThrLeuThrGlyAsnGlnMetL 284
 801 TTATAAACACTAGTACAGATCTAAGCTCACCCCTCAGTGAATCAGATG 850
 284 euLeuPheSerAsnAsnThrSerThrThrIleGlyIylAlaIleTYrVal 300
 851 TACTCTTCAGACAAATACATCGAACAGGGGAGGAGCTATCTATGTG 900
 301 LysLYSLeuGluLeuAlaSerGIyGlyLeuThrLeuPheSerArgAsnS 317
 901 AAAAGCTCGAAGTGGCTCCGGAGGACTTACCCTATTGAGTAGAATAG 950
 317 rValAsnGlyGlyThrAlaProLYSGlyGlyAlaIleAlaIleGluAspS 334
 951 TGTCAATGGAGGTACAGCTCTAAAGGTGAGCCATAGCTATCGAAGATA 1000
 334 eGlyGluLeuSerLeuSerAlaAspSerGIyAspIleValIlePheGly 350
 1001 GTGGGGAATTGAGTTATCCGCCGATAGTGCTGACATTTGCTTTTAGGG 1050
 351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
 1051 AATACAGTCACTTCTACTACTCTCGGAGCAATAGAGTAGTAGTGCAT 1100
 367 uGlyThrSerAlaIylSerThrAlaLeuArgSerAlaIleGlyArgAlaI 384
 1101 AGGACGAGGTGCAAGATGACAGCTTGGCTGTGCTGCTGAGTAGGCCA 1150
 384 leTYrPheTYrAspProIleThrThrGlySerSerThrThrValThrAsp 400
 1151 TCTACTTCTATGATCCCATACTACAGAGATCTTCCACACAGATTACAGAT 1200
 401 ValLeuLYSValAsnGluThrProAlaAspSerAlaLeuGlnTYrThrGI 417
 1201 GTCTTAAAGTTAATGAGATCCGGAGATTTCGCACTACAAATATACAGS 1250
 417 yAsnIleIlePheThrGlyIyluLYSLeuSerGIuThrGluAlaIleAspS 434
 1251 GAACATCATCTTCACAGAGAAAGTTATTCAGACAGAGGCCGAGATT 1300
 434 eLYSAsnLeuThrSerIysLeuLeuGlnProValThrLeuSerGIyGly 450
 1301 CTAAATCTTACTTCGAAAGCTACTACAGGCTTAACTTTTCAGAGAGG 1350

451 ThrLeuSerIleLYSHISGlyValThrLeuGlnThrGlnAlaPheThrGI 467
 1351 ACTATCTTAAACATGAGTAGCTCTGACAGCTCAGGACATTCACCTCA 1400
 467 ngIuAlaAspSerArgLeuGluMetAspValGIyThrThrLeuGluProA 484
 1401 ACAGGCGAGATTCTCGTCTCGAAATGACAGCTAGGAATCTATGAACTTG 1450
 484 laAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
 1451 CTGATACTACACCATTAACAAATTTGGTCATTACATCAGTTCTATAGAC 1500
 501 GlyAlaIylsValIylGlnIleGluThrIylsAlaThrSerIysAsnLeuTh 517
 1501 GGTGCAGAGAGCAAAATAGAACCAAAAGCTACGCTCAAAAATCTGAC 1550
 517 rLeuSerGIyThrIleThrLeuLeuAspProThrGIyThrPheTYrGluA 534
 1551 TTTATCTGGAACCATCACTTATTGAGCCGAGGGCAGCTTTTATGAAA 1600
 534 snHisSerLeuATGAsnProGlnSerTYrAspIleLeuGluLeuLYSAla 550
 1601 ATCATGTTTAAAGAAATCCTCAGTCTACAGACATCTTAGAGCTCAAAAGCT 1650
 551 SerGIyThrValThrSerThrAlaValThrProAspProIleMetGlyGI 567
 1651 TCTGGAGCTGTAAACAGCACCGAGTACTCCAGATCTATATATGGGTGA 1700
 567 uLYSPheHISTYrGIYrGlnGlyThrTrpGIyProIleValIyrGIyT 584
 1701 GAATATCCATTACGGCTATCAGGAACTTGGGGCCAAATGTTGGGGGA 1750
 584 hGlyIylAserThrThrAlaThrPheAsnTrpThrIylsThrGlyTYrIle 600
 1751 CAGGGGCTTCTACGACTGCACCTTCAACTGAGCTAAACCTGCTATAT 1800
 601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
 1801 CCGAATCCCGAGCGATACGGCTTTAGCCCTAATAGCTTATGGAATGC 1850
 617 aPheIleAspIleSerSerLeuHISTYrLeuMetGIuThrAlaAsnGlnG 634
 1851 ATTTATAGATATTAGCTCTCTCATTTATCTTATGAGACTGCAACAGAG 1900
 634 IYLeuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhehe 650
 1901 GGTTCAGGAGAGCCGTGCTTTTGGTGGTGAATATCTAATCTTCTTC 1950
 651 HISLYSAspSerThrLYSThrArgArgIylPheArgHISLeuSerGIyGI 667
 1951 CATAGGATAGTACAAAACACAGCGCGGTTTGCCTTTGAGTGGGGG 2000
 667 yTYrValIleGIyGlyAsnLeuHISThrCysSerAspLYSleuSerA 684
 2001 TTATGTCATAGAGAGAAACCTACATCTTGTTCAGATTAAGATTCTTAGTG 2050
 684 laAlaPheCysGlnLeuPheGlyArgAspArgAspTYrPheValAlaLYS 700
 2051 CTGCATTTTGTAGCTCTTGGAGAGATAGAGACTATTTGTAGCTAGAG 2100
 701 AsnGlnGlyThrValTYrGIyGlyThrLeuTYrTYrGlnHISAsnGluTh 717
 2101 AATCAAGGATAGCTACGAGAGACCTCTCATATACGACCAACGAAAC 2150
 717 rTYrIleSerLeuProCysLYSLeuArgProCysSerLeuSerTYrValP 734
 2151 CTATATCTCTCTCTTCCAAACTACGGCTTGTGCTGTTGCTATGTGC 2200
 734 roThrGluIleProValLeuPheSerGIyAsnLeuSerTYrThHISThr 750
 2201 CTACAGAGATTCCTGCTCTTTTCAGGAAACCTTAGCTACACCATAGC 2250

alignment_scores:

Quality: 4746.00 Length: 928
 Ratio: 5.114 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-428-122-2 x AE002235/rev ..

Align seg 1/1 to reverse of: AE002235 from: 1 to: 12127

```

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
  |||||
9110 ATGAAGTCTTCTTCCCAAGTTTGATTTCTACATTGCTATTCTTCC 9061

17 OLeuSerMetIleAlaThrGluThrValLeuAspSerSerAlaSerPheA 34
  |||||
9060 TTGTCATGATGCTTACCGAGACAGTTTGGATTCAAGTCGCAATTTCG 9011

34 SPGLYASNLYSASNGLYASNPHESERVALARGLUSERGLNGLNASPALA 50
  |||||
9010 ATGGGAATAAATAATGTAATTTTCAGTTCGTAGAGAGTCAGAGAATGCT 8961

51 GlyThrTrpTyrLeuPheLysGlyASNValThrLeuGluASNlleProG 67
  |||||
8960 GGAAGTACCTACCTATTAAAGGAATGTCACTCTAGAAAATATTCCNG 8911

67 YThrGlyThrAlaIleThrLysSerCysPheASNAsnThrLysGlyAspL 84
  |||||
8910 AACAGGCACACCAATCACAAAAAGCTGTTTACACACTAAGGCGCATT 8861

84 eutThrPheThrGlyASNGLYASNserLeuLeuPheGlnThrValAspAla 100
  |||||
8860 TGACTTTCACAGGTACGGGAAGCTCTATTGTTCCAAACGCGTAGATGCA 8811

101 GLYThrValAlaGlyAlaIleAlaValASNserSerValValAspLysSerTh 117
  |||||
8810 GGGAGCTAGACAGGGGCTGCTGTTTAACGACGCGTGGATTAATATCTAC 8761

117 rThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySers 134
  |||||
8760 CACGTTTAAGGTTTCTTCGCTATCTTTATTGCGCTCTCTGAGAGTT 8711

134 erIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSer 150
  |||||
8710 CGATAACTACCGCAAGAGACCCTGAGCTCTACGGGAGCTTAGT 8661

151 LeuThrLysASNValSerLeuLeuPheSerLysASNpheserThrAspAs 167
  |||||
8660 TTGCAC.AAAAATGTCAGTTTGCCTCTTACGCAAAAACCTTTTCAACGGATAA 8612

167 ngLYGLYAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMets 184
  |||||
8611 TGGCGGTGCTATCACCGCAAAAACCTTTTCATTACAGGAGCTACATGT 8562

184 erIleAlaLeuPheSerGluASNThrSerSerLysLysGlyGlyAlaIleGln 200
  |||||
8561 CAGCTCTGTTTCTGAAAATACCTCTCAAGAAAGCGGAGCCATTTCAG 8512

201 ThrSerAspAlaLeuThrIleThrGlyASNGLNGLYValSerPheSe 217
  |||||
8511 ACTTCGATGCCCTTACCATTTACTGGAACCAAGGAGGAAGTCTTTTTC 8462

217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluAlaSerV 234
  |||||
8461 TGACAAATACTTCTCGGATTTCTGGAGCTGCATTTTACAGAAAGCCTCGG 8412

234 alPhrIleSerAsnASNAlaLysValSerPheIleAspAsnLysValThr 250
  |||||
8411 TGACTATTCTTAATATGCTAAAGTTTCTTTATTGCAATAAAGTCACA 8362

251 GYLAlaSerSerSerThrThrGlyAspMetSerGlyGlyAlaIleCysAl 267
  |||||
8361 GGAAGCGAGCTCTCAACACAGGGGATATGTCAGAGAGTCTATCTGTGC 8312

```

```

267 aTYrLYsThrSerThrAspThrLysValThrLeuThrGlyASNGLMetL 284
  |||||
8311 TTATAAACTAGATACAGATACAGATACAGTCCACCTCAGGAAATCAGATGT 8262

284 euleuPheSerASNAsnThrSerThrThrAlaGlyGlyAlaIleTyrVal 300
  |||||
8261 TACTCTTCAGCACAAATACATCCGACAAACAGCGGAGAGAGCTATCTATGTG 8212

301 LysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerArgAsnSe 317
  |||||
8211 AAAAGCTCGAACGCTTCCTCCGAGAGCACTTACCCCTATTCTAGTAGAATAG 8162

317 rValASNGLYThrAlaProLysGlyGlyAlaIleAlaIleGlnAspS 334
  |||||
8161 TGTCAATGGAGGTACAGCTCTCTAAAGTGGAGGCATACGCTATCGAAGAA 8112

334 eRGlyGluLeuSerLeuSerAlaAspSerGlyAspIleValPheLysGly 350
  |||||
8111 GTGGGAATTGAGTTTATCCCGCGATAGTGTGACATTGCTTTTAAAGG 8062

351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
  |||||
8061 AATACAGTCACTTCTACTACTCTCGGAGGAATPAGAAGTAGATCGACTT 8012

367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAlaI 384
  |||||
8011 AGGACAGAGCGCAAGATGACAGACTTTCGCTTCTGCTGCTGATAGACCA 7962

384 leTYrPheTyrAspProIleThrThrGlySerSerThrThrValThrAsp 400
  |||||
7961 TCTACTTCTATGATCCCATTAACATACAGATCATCCACACAGTTACAGAT 7912

401 ValLeuLysValASNGLuThrProAlaAspSerAlaLeuGlnTyrThrCl 417
  |||||
7911 GTCTTAAAGTTAATAGAGACTCCGCGAGATTCTGCACTCAATATPACAAG 7862

417 yASNlleIlePheThrGlyGluLysLeuSerGluThrGluAlaAlaAspS 434
  |||||
7861 GAACATCATCTTACAGAGAAAGTTATCAGACAGACAGGCCCGCAGATT 7812

434 erLYsAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGlyGly 450
  |||||
7811 CTAAATATCTTACTTGGAAGCTACTACAGCTGTAACTCTTTCAGAGAGT 7762

451 ThrLeuSerLeuLysHisGlyValThrLeuGluThrGlnAlaPheThrCl 467
  |||||
7761 ACTCTATCTTTTAAACATGAGAGAGACTGTGCAGACTCAGGCAATTCACCTCA 7712

467 nglnAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGluProA 484
  |||||
7711 ACAGGAGATTCGTCGTCGAAATGAGAGCTAGGAAGTACTAGAACCTG 7662

484 lAspThrSerThrIleAsnAsnLeuValIleAsnLleSerSerIleAsp 500
  |||||
7661 CTGATACTAGCACCATTAACAAATTTGGTCATTACATCAGTTCTTATAGAC 7612

501 GLYAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuTh 517
  |||||
7611 GGTGCAAAAGAGCAAAAATACAAACCAAGCTACGTCAAAAAAATCTGAC 7562

517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGluA 534
  |||||
7561 TTTATCTGAAACCATCATCTTTATGTGACCCGAGCGGACGTTTATGAAA 7512

534 snHisSerLeuArgAsnProGlnSerTyrAspIleLeuGluLeuLysAla 550
  |||||
7511 ATCATAGTTTAAAGAAATCCTCAGCTCTAGACATCTTAGAGCTCAAAAGCT 7462

551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGlyG 567
  |||||
7461 TCTGAGACTGTACAAACAGCCAGTGTGACTCCAGATCTTATATATGGGTGA 7412

```

```

567 uTysepheniSTyrglyTyrglnGlyThrTrpGlyProIleValITrpGlyT 584
|||||
7411 GAAATTCCTTACGGCTATCAGGGAACCTGGGCCCAATTGTTGGGGA 7362
584 hGcIAlAseerThrThrAlaThrPheAsnTrpThrIysThrGlyTyrlle 600
|||||
7361 CAGGGGCTTCTACGACTGCACCTTCACTGCACCTTAACCTGGCTATATT 7312
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
|||||
7311 CCTAATCCGACGATACGGCTCTTATGCTTATAGCTTAATGGAATGC 7262
617 aheHeIleAspIleSerSerLeuHisTyrlleuMetGluThrAlaAsnGluG 634
|||||
7261 ATTTATAGATATTAGCTCTCTCATTTATCTTATGAGACTGCACCAAGAG 7212
634 IyLeuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhePhe 650
|||||
7211 GGTTCAGAGGACACCGCTGTTTGGTGTGCTGGATTACTACTTCTTC 7162
651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyG 667
|||||
7161 CATAGAGATAGTACAAAACACGACGGGTTCCGCATTTGAGTGGCGG 7112
667 YTyrValIleGlyGlyAsnLeuHisThrcysSerAspLysIleLeuSerA 684
|||||
7111 TTATGTCAATGAGGAAACCTACATCTGTCAGATAGATCTTAGTG 7062
684 IaAlaPheCysGlnLeuPheGlyArgAspArgAspTrpPheValAlaLys 700
|||||
7061 CTGCATTTGTCAGCTCTTGGAGAGATAGAGACTACTTGTAGCTAG 7012
701 AsnGlnGlyThrValTyrglyTyrglnLeuTyrryrglnHisAsnGluTh 717
|||||
7011 AATCAGAGGACTGTACGAGGAACCTCTATTTACCAACACCAAGAAAC 6962
717 rTyrlleSerLeuProCysLysLeuArgProCysSerLeuSerTyrryValp 734
|||||
6961 CATATCTCTCTTCTTGCAAACTACGGCTTGTGCTTATATGTTTC 6912
734 roThGluIleerovaAlaLeuPheSerGlyAsnLeuSerTyrrThrHisThr 750
|||||
6911 CTACAGAGATTCCTGTTCTTTCAGGAACCTTACGATACCACTATAGC 6862
751 AspAsnAspLeuLysThrLysTyrrThrTyrrProThValLysGlySe 767
|||||
6861 GATACGATCTAAAAACCAAGATATACACATATCTACTAGTTAAAGAGAG 6812
767 rTrpGlyAsnAspSerPheAlaLeuGluPheGlyGlyArgAlaProIleC 784
|||||
6811 CTGGGGGAATGATAGTTTCGTTTGAATTCGGTGAAGAGCTCGATTTT 6762
784 yLeuAspGluSerAlaLeuPheGluGlnTyrrMetProPheMetLysLeu 800
|||||
6761 GCTTGATGAAAGTCTCTATTGACATACATCCCTCTCATGAAATYG 6712
801 GlnPheValTyrrAlaHisGlnGlnGlyPheLysGlnGlnGlyThGluAl 817
|||||
6711 CAGTTTGTCTATGCAATCAGAGAGTTTAAAGAACAGGAACAGAAC 6662
817 aArgGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGlyI 834
|||||
6661 TCGTAATTTGGAAGTAGCGCTTGTGAATCTTGCTTACTTATCGGGA 6612
834 leArPheAspLysGluSerAspCysGlnAspAlaThrTyrrAsnLeuThr 850
|||||
6611 TCCGATTTGATAGGAATCAGACTCCCAAGATGCAACGACATCTAATCT 6562
851 leuGlyTyrrThrValAspLeuValArgSerAsnProAspCysThrThrTh 867
|||||
6561 CTGGTTATACTGCTGATCTTCTTGTAGTAAACCCGAGTACGACACAAC 6512
867 rLeuArgIleSerGlyAspSerTrpLysThrPheGlyThrAsnLeuAlaA 884

```

```

6511 |||||
ACTGGAATTCAGCGGATTCCTTGGAAAACTTCGTAACGAATTTGCCAA 6462
884 rGcIAlaLeuValleuArgAlaGlyAsnHisPheCysPheAsnSerAsn 900
|||||
6461 GACAAGCTTTAGTCTCTTCGTCAGAGGAACATTTTGTCTTAACTCAAT 6412
901 PheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSerArgAs 917
|||||
6411 TTGGAAGCTTTAGCCATTTCTTTGAATTCGTCGTCATCTCGCA 6362
917 nTyrrAsnValAspLeuGlyAlaLysTyrglnPhe 928
|||||
6361 TTCAATGTAGACTTAGAGCAAAATACCAATTC 6328

seq_name: gb_ba1:AE001587

seq documentation block:
LOCUS AE001587.16448 bp DNA BCT 08-MAR-1999
DEFINITION Chlamydia pneumoniae section 3 of 103 of the complete genome.
ACCESSION AE001587 AE001563
VERSION AE001587.1 GI:4376271
KEYWORDS
SOURCE
ORGANISM Chlamydia pneumoniae.
Chlamydia pneumoniae.
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
1 (bases 1 to 16448)
Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
Grimwood, J., Davis, R.W. and Stephens, R.S.
Comparative Genomes of Chlamydia pneumoniae and C. trachomatis
unpublished
2 (bases 1 to 16448)
Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
Grimwood, J., Davis, R.W. and Stephens, R.S.
Direct Submmission
Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
1..16448
/organism="Chlamydia pneumoniae"
/strain="GWL029"
/db_xref="taxon:83558"
172..2957
/note="Polymorphic Outer Membrane Protein G Family:
contains frame shift"
172..705
/genes="pmp_3.1"
/notes="Cpn0014"
615..2957
/genes="pmp_3.2"
/notes="Cpn0015"
3196..4871
/genes="pmp_4.1"
/notes="Cpn0016"
615..2957
/genes="pmp_5.1"
/notes="Cpn0018"
6302..9139
/genes="pmp_4.2"
/notes="Cpn0017"
6302..7786
/genes="pmp_5.1"
/notes="Cpn0018"
6302..9139
/notes="Polymorphic Outer Membrane Protein G Family:
contains frame shift"
787..9139
/genes="pmp_5.2"
/notes="Cpn0019"
complement(9380..11467)
/genes="Cpn0020"
complement(9380..11467)
/genes="Cpn0020"

```

[illegible]

CDS

CD5

BASE COUNT 4710 a 3330 c 3665 g 4743 t
ORIGIN

alignment_scores:
Quality: 4746.00 Length: 928
Ratio: 5.114 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-428-122-2 x AE001587 ..

Align seg 1/1 to: AE001587 from: 1 to: 16448

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
172 ATGAAGTCTTCTTCCCAAGTTGATTTTCATATTGCTATTTGCC 221

17 oleUserMetIleAlaThrGluThrValLeuAspSerSerAlaSerPheA 34
222 TTGTGCTATGATGCTACCGAGACAGTGGTGGTTGAAGTCGACAGTTGG 271

34 spoLysnLysAsnGlyAsnPhSerSerValArgGluSerGlnGluAspAla 50
272 ATGGAAATAAAATGGAAATTTTCAGTTGCTGAGAGTCAGGAAGTGTCT 321

51 GlyThrThrTyrLeuPheLysGlyAsnValThrLeuGluAsnIleProG 67
322 GGAAGTACCTACTATTAAAGGAATGTCACCTCTGAAATAATTCCTGG 371

67 ythrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspL 84
372 AACAGGCACAGCAATCAAAAAGCTGTTTAAACAACACTAAAGGGGAT 421

84 eauthrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla 100
422 TGACTTTCACAGGTAAAGGGAACTCTCTATGTTCCAAAGGTGATGCA 471

101 GlyThrValAlaGlyAlaAlaValAsnSerSerValValAspLysSerTh 117
472 GGAAGCTGACAGAGGCTGCTGTTAACAGAGCGTGATATAAATCTAC 521

117 rthrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySers 134
522 CACGTTATATAGGTTTCTCTGCTATCTTTATTGCGTCCCTCGAAGTT 571

134 ertLeuThrGlyLysGlyAlaValAserCysSerThrGlySerLeuSer 150
572 CGATTAATCAGGCAAAAGGACCGTTAGCTGCTCTACGGGTACTTGAGT 621

151 leuThrLysAsnValSerLeuLeuPheSerLysAsnPhSerThrAspAs 167
622 TTGAC. AAAAATGTCAGTTGCTCTTACGAAAACCTTTTCAACGGATA 670

167 nglyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThMetS 184
671 TGGCGGTCATACCGCAAAACTCTTTATTAACAGGACATCAATGT 720

```
184 eRAlaLeuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGln 200
|||||
721 CAGCGCTGTTTCTGAAAAATACCTCTCAAGAAAGGCGGAGCCATTGAG 770
201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValIleSerPheSe 217
|||||
771 ACTTCGAGTGCCTTACCATTTACTGAAACCAAGGGAAGTCTTTTTC 820
217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluAlaSerV 234
|||||
821 TGACAATATCTTCTCGATTCTGAGCTGCAATTTTACAGAACCTCGG 870
234 aThrIleSerAsnAsnAlaLysValIleSerPheIleAspAsnLysValIle 250
|||||
871 TGACTATTCTTAATATGCTAAAGTTCTTTATTGACAAATAGGTGACA 920
251 GlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaAlaIleCysAl 267
|||||
921 GGAGCGAGCTCCTCAACAACGGGGGATATGTCAGAGAGCTGATCTGTC 970
267 aTyrlLysThrSerThrAspThrLysValIleLeuThrGlnGlnMetL 284
|||||
971 TTATTAACACTAGTACAGATACATAAGTCACTCAGTGAATTCAGATGT 1020
284 euLeuPheSerAsnAsnThrSerThrThrAlaGlyAlaIleValIleThrVal 300
|||||
1021 TACTCTTACGACAAATATCATCGACACACGGGGAGGACTTCTATGTG 1070
301 LysLysLeuGlnLeuAlaSerGlyLysLeuThrLeuPheSerArgAsnSe 317
|||||
1071 AAAAGCTCGAACTGCTTCGGAGAGCTTACCTATTACAGTAGAAATAG 1120
317 rValAsnGlyLysThrAlaProLysGlyGlyAlaIleAlaIleGluAspS 334
|||||
1121 TGTCATGTGAGTACAGCTCTCTAAAGGTGAGCCCATGCTATCGAAGATA 1170
334 eRgLYgluLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
|||||
1171 GTGGGAAATTGAGTTATCCCGCGTACTGAGACATGTCCTTTTAGG 1220
351 AsnThrValIleThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
|||||
1221 AATACAGTCACTTCTACTACTCTCTGGAGCAATAGATAGATGACTT 1270
367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleAlaGlyAlaI 384
|||||
1271 AGGAACGAGTCAAAAGATGACAGCTTTCGCTCTGCTGTAGAGCCA 1320
384 lEtyrPheTyAspProIleThrThrGlySerSerThrThrValIleThrAsp 400
|||||
1321 TCTACTTCTATGATCCCATTACTACAGATCATCCAAACAGTTACAGAT 1370
401 ValLeuLysValAsnGlnLysThrProAlaAspSerAlaLeuGlnIleThrG 417
|||||
1371 GTCTTAAAGTTAATGAGACTCCGGCAGATTCTGCATACAAATATACAG 1420
417 yAsnIleIlePheThrGlyLysLeuSerGlnLysLeuSerGlnLysAlaIleAspS 434
|||||
1421 GAACATCTCTTCACAGAGAAAGTTATCAGAGACAGAGCGCGCAGATT 1470
434 eRlysAsnLeuThrSerLysLeuLeuGlnProValIleThrLeuSerGlyLys 450
|||||
1471 CTAATAATCTTACTCGAAGCTACTACAGCCGTGAACCTTTACAGAGGT 1520
451 ThrLeuSerLeuLysHisGlyValIleLeuGlnThrGlnAlaPheThrG 467
|||||
1521 ACTCTATCTTAAACATGAGTGAAGTCTGCACTGCAAGCAATTCATCA 1570
467 nGlnAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGlnProA 484
|||||
1571 ACAGGCAATTTCTGCTCGAAATGAGCTAGAGACTACTAGAACCTG 1620
484 lAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
|||||
1621 CTGATACTAGCACCATTAACATTTGGTCATTAACATCAGTCTATAGAC 1670
501 GlyAlaLysLysAlaLysIleGlyThrLysAlaThrSerLysAsnLeuTh 517
|||||
1671 GGTCAAAAGAGCAAAATAGAAACCAACGCTACGTCAAAATAATCTGAC 1720
517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrgL 534
|||||
1721 TTTATCTGGAACCATCATCTTATTGGACCCGACGGCAGCTTTATGAAA 1770
534 snHisSerLeuArgAsnProGlnSerTyArgIleLeuGlnLeuLysAla 550
|||||
1771 ATCATAGTTTAGAAATCTCATGCTTACGACATCTTAGAGCTAAAGCT 1820
551 SerGlyThrValIleThrSerThrAlaValIleProAspProIleMetGly 567
|||||
1821 TCTGGAAGTATACACAGACCGCAGTGAATCTCAGATCTTATATGGGTGA 1870
567 uLysPheHisTyrgLysTyrgLysGlnGlyThrTrpGlyProIleValTrpGly 584
|||||
1871 GAAATTCATTACGCTATCAGGGAAGTCTGGGGCCCAATTTGTTGGGGA 1920
584 hRgLYAlaSerThrThrAlaThrPheAsnTrpThrLysThrGlyTyrlle 600
|||||
1921 CAGGGGCTTCTACAGACTGCAACCTTCAACTGACTAAACGCGCTATATT 1970
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
|||||
1971 CCTAATCCCGAGCGTATCGGCTTGTAGTCCCTAATAGCTTATGGAATGC 2020
617 aPheIleAspIleSerSerLeuHisTyrlleMetGlnLysThrAlaAsnG 634
|||||
2021 ATTTATGATATATAGCTCTCTCCATTTATCTTATGAGACTGCAACAGAG 2070
634 lYleuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhePhe 650
|||||
2071 GGTTCAGAGGAGACCGTCTTTGGTGCTGTGATTAATCACTTCTTC 2120
651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyL 667
|||||
2121 CATAGAGTATGACAAAACACGAGCGGGTTCGCCATTGAGACTGGCGG 2170
667 yTyrlValIleGlyLysAsnLeuHisThrCysSerAspLysIleLeuSerA 684
|||||
2171 TTATGTCTATAGAGAAACCTACATACTTGTTCAGATTAAGATTCTTATG 2220
684 lAlaIlePheCysGlnLeuPheGlyArgAspArgAspTyrlPheValAlaLys 700
|||||
2221 CTGCATTTTGCACGCTCTTTCGAGAGATAGAGACTTGTGTAGCTAAG 2270
701 AsnGlnGlyThrValTyrgLysGlyLysLeuTyrlTyrgLysHisAsnG 717
|||||
2271 AATCAGGTACAGTCTACGGAGAACTCTATTATACAGCAACAGAAAC 2320
717 rTyrlIleSerLeuProCysLysLeuArgProCysSerSerLeuSerTyrlP 734
|||||
2321 CTATATCTCTCTCTTTCGAAACTACGCGCTTGTGTTGTTATATGTC 2370
734 rThrGlnIleProValLeuPheSerGlyAsnLeuSerTyrlHisThr 750
|||||
2371 CTACAGAGATCTCTGTTCTCTTTCAGGAAACCTTACTACACCCATACG 2420
751 AspAsnAspLeuLysThrLysTyrlThrThrTyrlProThrValLysGly 767
|||||
2421 GATAACGATCTGAACCAACAGTATACCAACATATCTCTGTAAAGGAG 2470
767 rTrpGlyAsnAspSerPheAlaLeuGluPheGlyLysArgAlaProIleC 784
|||||
2471 CTGGGGAGATATAGTTTGGTTAGAAATGCGTGAAGAGCTCGAATTT 2520
784 yLysLeuAspGluSerAlaLeuPheGlnGlnIleMetProPheMetLysLeu 800
|||||
```

/product="outer membrane protein 6"
 /protein_id="CAB37075.1"
 /db_xref="GI:4455891"
 /translation="MRSISGCPLEVFSLTLVSPDNLSTATITSLPDESEI
 ERSLNVAGDVLSLTGVDVTSISNDNLAKACNRVYSGSYTPAGNHGILYFN..SGCT
 TREKAVLACCCDPQATFAFSGFSTLSLFQSGDGLKEOCLSKNALMLNLVVFREON
 QSKIRKGALISGANTVIGVNDVSVEFYONMAFEGALHSGPLDIAVNOALIRPANOIC
 KNGSGALYSAGGIDIDONAVYLFRENEALPTAIGKGAVACCLPTSGSSSTPVIYES
 DNKGALVEPNHNSIGMDIYAARKLSISGGPTLEFNI SYANSQNLGALAIIDGGEI
 SLSEKGTITFQGNRTSLPFLNHIHLQNLKFLIKLORNCSTIEEYPIITSEADGVSQ
 LTNQSDKKNKTYCTTILFSGEKLANDRPPKSTIPONWLSGVYILKEGAVATSK
 FTQSGDNLVLVDGTLTASKEGAIITGLAIDIDSLSSTTAIVTANTKNOISVTD
 SIELISPGNAVIEDLRMRNSQTPPELLEFGAGSVTVIAGDELPSPIHFGQNMKL
 AMTQTNKVSQVGFEDMRKINYPKREPENLVNINMGAVNVSIMQVETHASLQWD
 RGLVLDIGNFEFVYASSEDNIRPHNSGGVLTAVNNNITPKRHTSNAFSOLFPRDKAY
 AVSNNEKRMVYSLYOYTTISLNI FPRYASRNPNVNGIISRRLQDPLMIFHLCAY
 GHAINDKKTQYANPMYKNSWRNMCNAIEGGSNPLPFEFGREAFGCAIFPMKLOLY
 AYQDQFKETTDAGRFPSNGSLTSTVPIGLREFELASQDVLVDPSFSYIPJDFKQD
 SCEALVYISGGSWLPAHVSRAHFAVSGSTGRHYFNFTYTELRCGSTIECPHARNTMI
 NCGSKFRP"
 498..3266
 /gene="omp5"
 3976..6870
 /gene="hb3"
 3976..6870
 /gene="hbl"
 /note="hypothetical"
 /codon_start=1
 /product="HB1 protein"
 /protein_id="CAB37076.1"
 /db_xref="GI:4455892"
 /translation="MRRLILAIYLALSLTLNGGVLTVAIPGSSVYSNPAGNG
 LADGEPQMLAGLIDVLRKREPVTLASVITTPETGSPRSISISGADITRSPLTL
 LCEGEPQMRKRLIAILVILFSLIASGVLTVAIPGLSSVYSNPAGNACALPCL
 VMLAGIDVLRKREVPVTLASVITTPETGSPRSIGISISGADITRSPLTPDEGPEI
 QSMKRLILAIYLIVFSLIASGVLTVAIPGLSSVYSNPAGNACALGVLTALAG
 IDVILKREVPVTPAPPEPVLDIDDEESIRLOQAEHALALPEMSAFEEYIKAVI
 VESHEENKSLPYDGHGEEKTRKQIQRVRESIKAMYPEPDIRIEPEEFPPLSAR
 KRLIDALTIVKRLITBQDERNNLRKAFSLTYODSIFKRTIDNEFKLANKFMILSS
 ILEDTITFENHEHGAKSLHKRNVLLEKVIYRSLQKSYRQIGMSKMKKLHGNPFF
 SCEDRKATIMEHEAMELSISRYKVALSDENVVDPSPDKWMDLSGJPCRALSES
 ISRDQOMKRAKHLKHOESLITQOARDRLTDOSSKNOKLEKAEQEYISMWERYKFEI
 ERVQRIATKQIKLVNPIEREETGTQETVPYQGTASGLDLDIIGLREVSREDDN
 QNOSCEVAVLSHEWEEMWEKQEGPKRKEFOOMGSLERFFHEIELELVLDQDYS
 KHLSPKVVNNKKEVQYAKPRLYKLESDEGLIAOTSASBLTQOELPIATGAELE
 KAVFGSICCLASAKAPYFEEDRFDQSDPQIALIRLOEKASLSPEETKRSNLE
 NDIAERRLLESKOTFERAGLVAREIAVSTSLDNLITPTEGTSSEKVVYSMTL
 NYNKEERRAKRTIVEMTQRYRDFEMALEAMQFNEALIQDELSTIQAPSE
 7150..8058
 /gene="hb2"
 7150..8058
 /gene="hb2"
 /note="hypothetical"
 /codon_start=1
 /transl_table=11
 /product="HB2 protein"
 /protein_id="CAB37077.1"
 /db_xref="GI:4455893"
 /translation="MOVLTSPOLPPPPRRSLGVSCKRSLSTITLVGLVLTILGII
 GLTATISGAGIGFALSGLGVYISLSEFLLYRREVPYTRSPFIRGVSVTPSEPALTE
 KAQKPEPKRLIDRLPKELDQDIDYIQGVFACLEKADPKIEDKRLILAEKAKLPVPE
 VYEKQEMSEEFDDIQVNLNEEAYVVEHCODPLENTAYLETFSSQELRDYICAVQGVLT
 GQARDRLKRSYKVEVMDRFMYRTKWSMASVMDHSYGVARELKKAGVLEESVYKTK
 LEKSYRDAFVECEKAKIQDGRFWL"
 8185..9492
 /gene="hb3"
 8185..9492
 /gene="hb3"
 /note="hypothetical"
 /codon_start=1
 /transl_table=11
 /product="HB3 protein"

```

/translation="MMLDRYADKFLIREKEEKMERHLEFATVNRASGHAVAKAKAA
FEKERSNENORXKVDKWKLSGLAEFRNOSRRARERLEQTLYPEVEEVEERLIER
ORFKKVLLENLYADIEKKYHGVREOEYHKAEYRENGEKVLSAEVSECCO
RLDCLETWMSKILTKAESVFEKMPDTEKLGKVLSDVNRLEIICEDAEEMIPFE
EITEMLRMVELPLPKNTFEKASLOVNSCKEMLAYEPOCKSPPTYSOEFLERLN
QDLQATINCQERLQGSDESQVTRCRDRLRQMKHEVOGLNFIINEELLVGAELEF
TOARLDVATVAPMEFYLOYHNIRKREVSQWMAKTEREIRQAFQGVKEDLLAED
TILKEEDYWLDRDMLRDERKKNRRLICNKIAAQORVYKGF"
9805.10698
/gene="hb4"
9805.10698
/gene="hb4"
/gene="hb4"
/note="hypothetical"
/codon_start=1
/translation="HB4 protein"
/product="HB4 protein"
/protein_id="CAB37079.1"
/db_xref="GI:4455895"
/translation="MLLLISGALFLTLGITGLSAISFGIGTSLALGGLVMTISLLC
LIVKREIPVPEEIPPEGVSLAPSEPALQAOKTLAQPKELDQDLDIOEYFACL
KLVDSYERSFNLDAKKELVDFVVEDLSEIFELRQIYAOEGDNLNLINGRSL
MTAESSELDLPHVSKRGLYPSGDVAREGLKRSKIVARLMSLHCEIKRVADFR
NSYAMAEKAFKALGALAESVYRSLQSYRDKFLSEHRAKIPNNGHITMLRDKAGSC
AEKLGMPRVGKRLGKQSG"
10776.12119
/gene="hb5"
10776.12119
/gene="hb5"
/note="hypothetical"
/codon_start=1
/translation="HB5 protein"
/product="HB5 protein"
/protein_id="CAB37080.1"
/db_xref="GI:4455896"
/translation="MDEIFPHELYEKTFLKRLHRGALAKTTEKKRSKKNLOAVEE
ANARLKYVDWYDOEFKAGERLEKHLALYPRVSVIRNKIOETRSNKEAYEIE
ENRCVREOEDWYKEEKREAFERENGKILSPEDLESLEDFDGLKFKSKMEL
EGHILKQKRAVEKRIISDNESREIYFEDVKEKPECEIEIKTLRAELPLPLT
KKAFEKACSOYNSCAEMLEKVPYCKESLAVYSKEVLSDLEDLRAAYTECKRQF
DSGLESEVRACQRLREIOEFOTGLVEKELVSSRLNTEDCVSGVKEKAP
GKKFYQYDEIVRVRQSRMTSERLREGVACNMLKAGLSEEDKVLKEEYMLY
REERKNKEKLVGTAKIVAQORIQEFOPSDIVSSNKKVLMOKARFLFRBDHS"
12362.12987
/gene="hb6"
12362.12987
/gene="hb6"
/note="hypothetical"
/codon_start=1
/translation="HB6 protein"
/product="HB6 protein"
/protein_id="CAB37081.1"
/db_xref="GI:4455897"
/translation="MTAAPALIHVSPTPEETKVIIPKDSKRALGITLIVGILLV
CGAIVSGVIGSALIVGLIGTISLIGVFLGLILLRLRELTLEQIEAKOLAE
TFDELKELMYTOSTEKSEKIEGSRVNOGFLNATOKIILDESISITSEPPDL
RQLEDEKIEFILSGERLLEFIANLEFKQGRDVLNLGNLADIAVYGNPNYKAMVIE
KAKAVHEFIVLTLMARELEFF"
13014.14585
/gene="hb7"
13014.14585
/gene="hb7"
/note="hypothetical"
/codon_start=1
/translation="HB7 protein"
/product="HB7 protein"
/protein_id="CAB37082.1"
/db_xref="GI:4455898"
/translation="MYOENLRLERLLYNSVOKSVADRLFSYKTKMVDPTPLIPWE
DKKCAAEKAPLEOKILIDYKSTFWLNEDEINLNDPMSGLNTVTRKRYFOVD
DSERMHKVLQKLENDYKELLESSESTEARKNLSDLVLEDAKTLFFLAKQEE
VETRVKDLRARVGTVPKODTEAKKRVLEASLETFLDSIESELVQCELDODIYWE

```

```

QDVKDLARTQLEBEDIEAKREBAEDRLSNRLKSKTMTLDRAKWHIENASITW
WTSQLEMKDMKARLKLKEDITSVPEIDEITCISLEPLTLTRELTKSYLKFKI
CSETLLKMTSVENNIYVOEYVQNLQKQISORRGKODOPANIEQVALQK
RRELQNEIEIGFNFMEEDFKAANDLYIRSTABOKNFDVCELPFRYHEVYK
LLEIMVNCDSYRDAKAKKCSRLDERKELQNEIKKEERYOKKQRRHADRSHITTYOK
LLEAELELTKKI"
14954.17479
/gene="omp7"
14954.17479
/gene="omp7"
/note="putative"
/codon_start=1
/translation="outer membrane protein 7"
/product="outer membrane protein 7"
/protein_id="CAB37083.1"
/db_xref="GI:4455899"
/translation="MKIPRFLILSVPLTSMNLGAATTEELASNSNFDGTTSTS
FSKTSSTARDGTNYVRKDSVLENNPKTGETQSTGCFKNDAAGDLNFIKGGFSFTFS

```

alignment_scores:

Quality	4746.00	Length:	928
Ratio:	5.114	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

us-09-428-122-2 x CPN133035 ..

Align seg 1/1 to: CPN133035 from: 1 to: 26920

```

1 MetlyssSerPheProlyPheValPheSerThrPheAlaIlePhePr 17
|||||
17762 ATGANGTCTTCTTCCCAAGTTGTATTTCTCATTTGCTATTTTCCC 17811
17812 TTTCGCTATGATTCTCTACCGACAGTTTGGATTCAAGTCGAGTTTGC 17861
17 OLeuSerMetIleAlaThrGluThrValLeuAspSerSerAlaSerPhe 34
|||||
17862 ATGGGAATAAAGTATGATTTTTCAGTCTGTGAGAGTCAGGAAGATGCT 17911
51 GLyThrThrTyLeuPheLeuGlyAsnValThrLeuGluAsnIleProG 67
|||||
17912 GGAAGTACCTACGTAATTTAAGGAAATGTCCTCTAGAAATATCTCGG 17961
67 YThrGlyThrAlaIleThrLySerCysPheAsnAsnThrLySGLysPL 84
|||||
17962 AACAGGCACACCAATCACAAAAGCTGTTTAAACACACTAAGGCGCAT 18011
84 eutrrPheThrGlyAsnGlyAsnSerLeuPheGlnThrValAspAla 100
|||||
18012 TGACTTTCACAGGTACGGGACTCTCTATTGTTCCAAACGCTGAGATGCA 18061
101 GLyThrValAlaGlyAlaValAsnSerSerValValAspLySerThr 117
|||||
18062 GGGACTGTAGCAGGGGCGCTGTACACGACGCTGTGAGATTAATCTAC 18111
117 rThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySer 134
|||||
18112 CACGTTTATAGGTTTCTTCCTGCTATTATGCGCTCTCCCTGGAAGTT 18161
134 erIleThrThrGlyLyGlyAlaValSerCysSerThrGlySerLeuSer 150
|||||
18162 CGATTAACACCGGCAAGAGACCTTACCTCTCTACGGGTAGCTTGAAGT 18211
151 LeuThrLyAsnValSerLeuLeuPheSerLyAsnPheSerThrAspAs 167
|||||
18212 TTGAC AAAAATGTCAGTTGCTCTCTCAGCAAAAACCTTTCAACGATAA 18260
167 nGlyGlyAlaIleThrAlaLyThrLeuSerLeuThrGlyThrThrMet 184
|||||
18261 TGGCGGTGCTATACCGCAAAAACCTTTTCAATTAACAGGACTACAAATGT 18310

```

```

184  eralaleuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGln 200
    |||
18311 CAGCTGTGTTTGTGAAATACCTCTCAAGAGAGCGGCATTCAG 18360
201  ThrSeraspAlaLeuThrIleThrGlyAsnGlnGlyIuValSerPhe 217
    |||
18361 ACTTCGAGTCCCTTACCATTACTGGAACCAAGGGAGTCTTTTTC 18410
217  rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluIleSerV 234
    |||
18411 TGACAATACCTCTTCGAGATCTCGAGCTCGCAATTTTACAGAGCTTCG 18460
234  aThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysValThr 250
    |||
18461 TGACTATTTCATTAATGCTAAAGTTTCCTTATTGACATTAAGTCA 18510
251  GlyIaSerSerSerThrThrGlyAspMetSerGlyGlyAlaIleGlySAI 267
    |||
18511 GGAGCGAGCTCTCAACACAGGGGATATGTCAGAGGTGCTATCTGTC 18560
267  aTyIleThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
    |||
18561 TTATAAACTGTACAGATCTAGTACAGTCACTCGAGAAATCAGATGT 18610
284  euLeuPheSerAsnAsnThrSerThrThrAlaGlyGlyAlaIleTyVal 300
    |||
18611 TACTCTCAGCAACATATACATCGACACAGCGGAGAGCTATCTATGTG 18660
301  LysIleLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerArgAsnSe 317
    |||
18661 AAAAAGCTCGAGCTGCTCCGGAGGACTTACCTATTTCAGTAAAGATAG 18710
317  rValAsnGlyGlyThrAlaProLysGlyGlyAlaIleAlaIleGluAsp 334
    |||
18711 TGTCATGAGAGTACAGCTCTCTAAGGTGAGCCATAGCTATGCAAGTA 18760
334  ergIleGluLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
    |||
18761 GTGGGGATTGAGTTTATCCGCCATAGTGTGACATTGCTTTTATAGG 18810
351  AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspIe 367
    |||
18811 AATACATCTCCTTACTACTCTCGGAGCAGATGGAAGTATGATTCGACTT 18860
367  uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAlaI 384
    |||
18861 AGGAACGAGTGCAAGATGACAGCTTGCGTTGCTGCTGTAGAGCA 18910
384  leTyrrPheTyrrAspProIleThrThrGlySerSerThrThrValThrAsp 400
    |||
18911 TCTACTCTCTATGATCCCATTAACCTACAGATCATCCACACAGTTACGAT 18960
401  ValLeuLysValAsnGluThrProAlaAspSerAlaLeuGlnTyrrThrG 417
    |||
18961 GTCTTAAAGTTATGAGACTCCGGCAGATTCTGACTACATATTAACAG 19010
417  yAsnIleIlePheThrGlyGlyLeuLysLeuSerGluThrGluAlaIleAsp 434
    |||
19011 GAACATCATCTTCACAGAGAAAAGTTATCAGAGACAGAGCCGACAGATT 19060
434  eIlysAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGlyGly 450
    |||
19061 CTTAAAAATCTTACTCGAAGCTACTACAGCTGTAACTCTTTCGGAGGT 19110
451  ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
    |||
19111 ACTCTATCTTTAAACATGGAGTACTCTGACAGCTCAGGCATTCACTCA 19160
467  nGlnAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGluProA 484
    |||
19161 ACAGGCGAGATTCTGCTCGTCAAAATGACGTAGGAACCTACTAGAACCTG 19210
484  laAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500

```

```

19211 CTGATCTACACCATTAACAATTTGGTCTATTAACATCACTCTATAGAC 19260
501  GlyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuTh 517
    |||
19261 GGTGCAAGAGAGCGAAAATATGAACCAAGCTACGTCAAAATAATCTGAC 19310
517  rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyGluA 534
    |||
19311 TTTATCTGGAACCATCTCTTATTGGACCCGACGGCAGCTTTTATGAAA 19360
534  snHisSerLeuArgAsnProGlnSerTyrrAspIleLeuGluLeuLysAla 550
    |||
19361 ATCATAGTATTAGAAATCTCAGCTCAGACATCTTAGAGCTCAAGACT 19410
551  SerGlyThrValThrSerThrAlaValThrProAspProIleMetGly 567
    |||
19411 TCTGGAACCTGTACAAACACCGCAGTGCATCCAGATCCATAATGGGTGA 19460
567  uLysPheHisTyrrGlyTyrrGlnGlyThrTyrrGlyProIleValTyrrGly 584
    |||
19461 GAATTCATTTACGGCTATCAGGGAAGTTCGGCCCATTTGTTGGGGA 19510
584  hrcGlyAlaSerThrThrAlaThrPheAsnTrpThrLysThrGlyTyrrIle 600
    |||
19511 CAGGGCTTCTACGACATCGCAACCTCACTGACATCAAACTGCTATATT 19560
601  ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
    |||
19561 CCTAATCCGAGCGGTATCGGCTCTTATGCTCTTAATAGCTTATGGAATGC 19610
617  aPheIleAspIleSerSerLeuHisTyrrLeuMetGluThrAlaAsnGlu 634
    |||
19611 ATTTATAGATATTAGCTCTCTCATTTATCTTATGAGACATCGCAAGAG 19660
634  lyleuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhePhe 650
    |||
19661 GGTTCAGGAGACCGGTCTTTTGGTGTGCTGATTAATCACTTCTTC 19710
651  HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGly 667
    |||
19711 CATRAGGATAGTCAAAAACACGACGCGGTTTCCCATTTGCTGTGCGG 19760
667  yTyrrValIleGlyGlyAsnLeuHisThrCysSerAspLysIleLeuSerA 684
    |||
19761 TTATGTCATGAGGAACCTACTACTGTTCAGATTAAGATCTTAGAG 19810
684  laAlaPheCysGlnLeuPheGlyArgAspArgAspTyrrPheValAlaLys 700
    |||
19811 CTGCATTTTGTCAAGCTCTTGGAAAGATAGAGACTTGTAGCTAAG 19860
701  AsnGlnGlyThrValTyrrGlyGlyThrLeuTyrrTyrrGlnHisAsnGlu 717
    |||
19861 AATCAAGGTACACTACGAGAGAACTCTTATTAACAGACACAGCAAAAC 19910
717  rTyrrIleSerLeuProCysLysLeuArgProCysSerLeuSerTyrrValP 734
    |||
19911 CTATATCTCTCTCTCTGCAAACTACGGCTTGTGTGCTATAGTTC 19960
734  roThrGluIleProValIleuPheSerGlyAsnLeuSerTyrrHisThr 750
    |||
19961 CTACAGAGATTCCTGTTCTCTTTCAGAAACCTTAGTACACCCCTAAG 20010
751  AspAsnAspLeuLysThrLysTyrrThrThrTyrrProThrValLysGlySe 767
    |||
20011 GATACAGATCTGAAACCAAGTATACACATATCCACTGTTAAAGAGAG 20060
767  rTyrrGlyAsnAspSerPheAlaLeuGluPheGlyGlyArgAlaProIleC 784
    |||
20061 CTGGGGAATGATAGATTTCGCTTAAATTCGATGGAAGAGCTCCGATTT 20110
784  ysLeuAspGluSerAlaLeuPheGluGlnTyrrMetProPheMetLysLeu 800
    |||

```



```
20111 GCTTAGATGAAAGTGTCTATTGACAGTACATGCCCTTCATGAAATTG 20160
      801 GlnpheValTyrAlaHisGlnGluGlyPheGlyGlnGlnGlyThrGluAl 817
      20161 CAGTTGTCTATGCACATCAGGAAGCTTTTAAAGACAGGAGACGAGAC 20210
      817 AARGGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGlyI 834
      20211 TCGTGAAATTTGGAACTACCGCTCTGTGAATCTTCCCTTACCTACCGGA 20260
      834 LeArgPheAspGlySerGluSerAspCysGlnAspAlaThrTyrAsnLeuThr 850
      20261 TCCGATTTGATAGCAATCAGACTGCTCCAGATGACAGCTTACATTTAAGT 20310
      851 LeuGlyTyrThrValAspLeuValArgSerAsnProAspCysThrThrTrh 867
      20311 CTGGGTATACTGTGATCTTGTTCGTAGTAAACCCGACTGTAGACAGAAC 20360
      867 rLeuArgIleSerGlyAspSerTrpPlySerThrPheGlyThrAsnLeuAla 884
      20361 ACTGCGAATTTAGCGGTGATCTTGGAAGAACTTCGTCGATGGAATTTGGCAA 20410
      884 rGlnAlaLeuValLeuArgAlaGlyAsnHisPheCysPheAsnSerAsn 900
      20411 GACAAAGCTTTAGTCTCTGTCGACAGGACCAATTTTGTCTTAACCAAT 20460
      901 PheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSerArgAs 917
      20461 TTGTGAAGCCTTTAGCCAAATTTCTTTGAATTCGCTGGTCATCTCGCNA 20510
      917 nTyrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
      20511 TTACATATTAACCTTAGACGCAAAATACCAATTC 20544
seq_name: gb_bai:AP002545
seq_documentation_block:
LOCUS AP002545 300050 bp DNA BCT 07-JUL-2000
DEFINITION Chlamydomophila pneumoniae genomic DNA, complete sequence, section 1/4.
ACCESSION AP002545 AB033780 AB033781 AB033792 AB033793 AB033794 AB033795
AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073
AB036074 AB036075 AB036076 AB036077 AB036078 AB038345 AB038346
AB038347 BA0000008
AP002545.1 GI:8978373
VERSION
KEYWORDS
SOURCE Chlamydomophila pneumoniae (strain:J138) DNA.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
AUTHORS Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
Shiba,T., Ishii,K., Hattori,M., Kunari,S. and Nakazawa,T.
TITLE Comparison of whole genome sequences of chlamydia pneumoniae J138
from Japan and CW029 from USA
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)
20330349
2 (bases 1 to 300050)
REFERENCE
AUTHORS Shirai,M.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.
Mutsunori Shirai, Yamaguchi University School of Medicine,
Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi
755-8505, Japan (E-mail:shirai@po.cc.yamaguchi-u.ac.jp,
Tel:81-836-22-2227, Fax:81-836-22-2415)
COMMENT AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)
AB038345-AB038347: Submitted (14-Feb-2000)
AB036071-AB036078: Submitted (18-Dec-2000).
FEATURES
source
1. 300050
/organism="Chlamydomophila pneumoniae"
/strain="J138"
/db_xref="taxon:83558"
/note="synonym:Chlamydia pneumoniae"
```

```
gene complement(1..282)
CDS /gene="Cpf0001"
complement(1..282)
/gene="Cpf0001"
/codon_start=1
/transl_table=11
/product="Ct001 hypothetical protein"
/protein_id="BA98211.1"
/db_xref="GI:8978374"
/translacion="MLGKIRGLSLIYLALANGLIGITPHNKNIITAKLCGGVSTP
ATQIYIIGINGVCLISFCFCFSKSRHSHGSDCSSGGCASHSDKN"
573..878
/gene="gacC"
573..878
/gene="gacC"
/codon_start=1
/transl_table=11
/product="Glu-tRNA Gln amidotransferase (C subunit)"
/db_xref="GI:8978375"
/translacion="MEQPHLDREILILLAKASALQISELLIOEYOTSASVITSKEA
LAIEIDDADESCLEFMHVNVNEDREDSVTSDFNREEFLRNVPESLGLVYPAVVK"
895..2373
/gene="gata"
895..2373
/gene="gata"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (A subunit)"
/protein_id="BA98213.1"
/db_xref="GI:8978376"
/translacion="MYRSALFLAAVTLGELTAGVQVQHEFHREAEAGVGARISL
KEQALQEQBELIDKRSRGEPLGAGVPGIKDNIHVGTIKTTASVLENYQPPD
ATVVERIKKEDGIIILKLNMEFAMGSTTLYSAFPHTPMDLTSVPGSSGGSAAV
SARFCPVALGSDTGGSIKOPAFCGVVEGFSKPSYSGVAVSRGLVAFNSLDOIGPLANTV
EDVALMPVEGSRDPRDARSREFRDSFMSKLTSTVPKIVIGVPRFLEGLRDIRENF
FSSLAFPFGEETHLVDELIDLSHAVSIYIILASAEATNLARFEGVGRSPQAT
ISOLDLSRGEGFEKVMRRLILGNYVISAEBQNYVKKATVAPVKIKYKARTAEKC
ELIAMPVCSRPAPFEIGELIDPVTLLKQDITVYANMLATLPALVPSGFSKEGLPLGLQ
IIGQGDQDQVQVGVSFQEHQIKQLFSKRYAKSVLGGOS"
2370..3836
/gene="gatB"
2370..3836
/gene="gatB"
/codon_start=1
/transl_table=11
/product="(Pct112) Gln tRNA Gln amidotransferase (B
subunit)"
/protein_id="BA98214.1"
/db_xref="GI:8978377"
/translacion="NSAYYADWESYIGLEVHVELNTASKLESSALNREGDEPNRIST
VCTGPGSLPYLNQSAVERKAVLFGVAGEISLSRFDKSYFYDPSRNPQIQFEP
PIIRGRIRKAIYQGEERYELAQTHIEDDAQMLKHFGEAGVDYRAGVPELIVSKP
CMFCEPDAAVATSLVSLDYIGISDCMNEGSIKFDVNAVSVRPGSPDELNRKEIKN
NMSFAMQALAEAKROIDEYLNQNPDKPLVIPAAYNVADPEKKTVLNRKESAE
DYKVFPEPDPLTQLESYIERIKTLPDLKYDHRARYIOEYGSSEDIASLISXNKI
ATFEVACKDCKNRSLSNMYSVEGRCIKLGYKLPSSGIFPEBVADLVNADQGI
TGKIAKETADLMESPGKRNPEILKEPELLPMSDEGLQKTLAEVVLANPESTIVDK
NGKTKALGFLVQIMKRTAGKAPKRVNELLLELDKG"
4127..6895
/gene="pmp_1"
4127..6895
/gene="pmp_1"
/codon_start=1
/transl_table=11
/product="polymorphic outer membrane protein G family"
/protein_id="BA98215.1"
/db_xref="GI:8978378"
/translacion="MRFSLGCPPLVFSFTLSVPTLSATTISLTPEDSRHGSQNA
RSYVVOAGDYSLNGDYISNVNSALNKAFCNVTSSYFAGCHHGLVNNINSSGT
TRGAVLCCQDPQATARSGETSFTSFTOSPEDIKQGLCKRNALMLNNVVFQON
OSKTKGAIISGANVITVGNVDSVFYQNAATFGGAIHSSGQLQIAVNOAEIRFQNTA
KNGSGALYSQGDIDIDONAVVLFRENEALLTAIGKGAGVCLPTSGSFTVPVPTVS
```

```

DNKQVFNHNSIMGGAIAIYARKLSSGGPTLFINNISYANSONLGAIAIDTGEI
SLSEKGTITFOGRTSLPFLNGHLLONAKFKLOARNGXTEFPDITSEDSSTQ
LNTGDPKNEKYGTITLIFSEKSLANDPRPKSTIPONVALSGITLKGAEVTVSK
FTOSPGSHVLIDGTCLIKLASKEDIALTGLAIDIDSSSSSTAIVIANANKOISVD
SIELISTGNAVEDLRMRNSQTEPLLSLEPGAGSVTVAGDFLPVSPHAFGOMKL
AMWGTGKVGEEFMDKINRKPREKBNLVPNTLIMGNAVDRSLMOVQEHASLQTD
RCGLMIDIGNEFHVASSEDNIYRHNNGSVYLVSNNETPKHYISMAFSQLEFRDXY
AVSNNERYMGLSYLOYTSLGNIFRYASRNPNVANGILSRRELONPLMIFRCLAY
GHATNDKTDYANFPMVKNSMRNNCAIECGSGMPLFPGALFQCALPDKMQLQVY
AYOGDFKETTADRRFNSGLSTISVPLGIRFKALSDQVLYDEFSFTIPDIFKDP
SCRAALVTSQSWLVPAAHVSRHAFVSGTGRYHFNDYTELLCRGSIIEGRPHARNI
NCGSKERE"
/gene="CPj0006"
/complement (7138. .7293)
/gene="CPj0006"
/complement (7138. .7293)
/gene="CPj0006"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98216.1"
/db_xref="GI:8978379"
/translation="MQEPLRSALLERLSEWLVLLGVSPETTRSTPEKANQLPKDR
NRLRESL"
/gene="CPj0007"
/6605. .10499
/gene="CPj0007"
/7605. .10499
/gene="CPj0007"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98217.1"
/db_xref="GI:8978380"
/translation="MRKRLIAIVLIALSIILAGSVLLTVAIPGLSVISPPAGM
AALGVMLAGIDVLKKREYIVLASVTTGATGTPGSPGSSISGASTISLPTVL
LDEGHQSMRKLRIALIVLIVESIILAGVLLTVAISGLSVISPPARMACALGC
VMAIGIDVLKKREYIVLASVTTGATGTPGSPGSSISGASTISLPTVL
QSMRKLRIALIVLIVESIILAGSVLLTVAIPGLSVISPAEWGACALGVMLAG
IDVLLKREYIVLPAPLPEEVVIDIDIESIRLOQEAALARLPEEMSAFEGYIKY
VSHLENNKSLPDGHELEKTHQIRVRSLSLKAQVPEFLIRLFEDEEFPFYSAR
KRLLDATTLEVERKILTEOLENNLRKAFSYODSIFKKIIDNEKLAAMKMLIS
ICRFTIIFENHEGVAKSLHKNNAVLEKVIYRSLOKSYRDIQMSAKKKILHGNPF
SLEDNKTIKMEAEMLSELSYRKVFALSDENVDPSPDKKMDLSIPDRDLSE
ISRDQWOKKAHLKHESLYTOARDRLTOOSKENOKLEKAEQETISWEVKFEI
EYQERTIOALQKLPNLEEBEETGQETVPTVOGTASOLDILGRIEWYSKEDN
QWQESCVKQVDRSHVEWSEWKOEGPKRKEQDOMGSLERFTEHIELEVLQKDS
KHLSTFKYNNKKEVOYAKFRLKLVESDEGLAQTESASLTLQEEPLIATRGALE
KAVFGSLCCALASKAKPYFEDPREDDQTLRALTLIOAKASLEBEIKRSENL
NDIAEERRLKESKOFERAGLGVLRETAVESTYDLRLSTNWEQTPPESEKYFPMYL
NYNNEKRRAKRIYEMTORRYDFKMALEAMQFNEDALLOEELSTIOAFSE"
/10975. .11668
/gene="CPj0008"
/10975. .11668
/gene="CPj0008"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98218.1"
/db_xref="GI:8978381"
/translation="MISGLFLVLRREVTVRSEIIPRGVSVTPSEEPALKEAKEPE
TKKILDPKELDQDLYIOEYFACLERLKDQKYPVLEAKKILVPPVVEKMDM
SEFLIDQVLINEAVYVHCODPLENINAEIFSSSELDIYCAVCGILPFGDAADR
LKRSYKEVMDRMRYTKWSWEASVMDHSYGARLFRKAVGLEESVYKILFKSYRD
AFYECEKAKIQDGRFKWL"
/11815. .13122
/gene="CPj0009"
/11815. .13122
/gene="CPj0009"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98219.1"
/db_xref="GI:8978382"
/translation="MMWDRYADKFLIREKEKMERHELFHATMWRKASGHAYAKAKA

```

```

FEKESNENOKRVKVEWLSKGLAEFNQDSRRARERLRELQTLYPEVSEVLER
QTKRVNLENIYADLEKRYHHCVRQEHYMKVEKKEVENEKVLASAEVSECLQ
RLEDDLETWSKRLTAESEVEMKFDATKGNVLSQVTRLEILCEDABEMFRIE
RIEMTLRWEVLPLEMKNTPEKASLOYNCKEMAKVPECKESPTVRSQERLELN
OLOJATYNNCOERLOGFSDLSKVTCDHILREOMKHEVGNLFINELLWGAELF
TOARLDATVATPYMEFYLOYHNIRREKRSQMSKTERYRRIROAFQGVKEDLLAE
TLKREKEDVLRDDMLLDBERKNRRLICRKIAAQORVAGF"
/gene="CPj0010"
/13434. .15758
/gene="CPj0010"
/13434. .15758
/gene="CPj0010"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98220.1"
/db_xref="GI:8978383"
/translation="MLLISGALFLTLGIPGLSAISFGLGIGLSALGVLMISGLC
LLVRKEIPTVPEEIPESVSLAPSEEPALQAQKTLAQPRELDQDPTDIOEVFACLR

```

alignment_scores:

Quality:	4746.00	Length:	928
Ratio:	5.114	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-428-122-2 x AP002545 ..

Align seg 1/1 to: AP002545 from: 1 to: 300050

```

1 MetLysSerSerPheProLysPheValPheSerThPheAlaIlePhePr 17
|||||
21400 ATGAAGCTCTCTTCCCAAGTTGATTTTGCATTTGCTATTTTCC 21449
17 cLeuSerMetIleAlaThrGluThrValLeuAspSerAlaSerPhea 34
|||||
21450 TTGTGTATGATGATTCACCGACAGACTTTTGATTCACACTCGAGATTTCG 21499
34 spcLysnLysAsnGlyAsnPheSerValArgGluSerGlnGluAspAla 50
|||||
21500 ATGCAATATAAATGCTATTTTTCGTCGTCGACAGATAGCAAGATGCT 21549
21500 GlyThrThrTyLeuPheLysGlyAsnValThrLeuGluAsnIleProgl 67
|||||
21550 GGAACCTACCTACCTATTTAAGGAAATGCACCTAGAAAATATTCCTGG 21599
67 yThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspL 84
|||||
21600 AACAGGACAGCAATACAAAAGCTGTTTAAACAACACTAAGGCGCAT 21649
84 eutThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla 100
|||||
21650 TAGACTTTCACAGGTAAAGGGAACCTCTATTTGTTCCAAACGGTGCATGCA 21699
101 GlyThrValAlaGlyAlaAlaValAsnSerSerValValAspLysSerTh 117
|||||
21700 GGAACCTGATGACAGGGCTGCTGTTAAACAGCAGCGTGTAGATAAATCTAC 21749
117 thrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySerS 134
|||||
21750 CACGTTTAAAGGGTTTCTCGCTATCTTTATATGCGTCTCTGGAAGATT 21799
134 erIleThrThrGlyLysGlyAlaValAserCysSerThGlySerLeuSer 150
|||||
21800 CGATACACTACCGGCAAGAGCGCTTACTCTCTACGCGGTAGCTTGACT 21849
151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPheSerThrAspAs 167
|||||
21850 TTGAC.AAAAATGTCAGTTGCTTTCAGCAAAAACTTTTCAAGGATAA 21898
167 nGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMetS 184
|||||
21899 TGGCGGTGCTATCACCGCAAAACTTTTCATTACACAGGAGCTACCAATGT 21948

```

```
184  erAlaLeuPheSerGluAsnThrSerSerLysLysGlyAlaIleGln 200
|||||
21949  CAGCTGCTTTCTGAAATACCTCTCAAGAAAGCGGAGCATTCAG 21998
201  ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerPheS 217
|||||
21999  ACTTCGAGTCCCTTACCTACTGGAACCAAGGAGAGTCTCTTTTC 22048
217  rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluLase 234
|||||
22049  TGACAATAATCTTCTCGATCTCGAGCTGCAATTTTACAGAACGCTCG 22098
234  aThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysValThr 250
|||||
22099  TGACTATTCTTAATAATGCTAAAGTTCTCTTATGACAATAAGCTCA 22148
251  GlyAlaSerSerSerThrGlyAspMetSerGlyAlaIleCysAl 267
|||||
22149  GGAGCGAGCTCTCAACACGGGGATATGTCAGAGAGTCTATCTGTGC 22198
267  aTyrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMet 284
|||||
22199  TTATAAAACTAGTACAGATCTAAGGTCACCTCAGCTGAAATCAGATGT 22248
284  euLeuPheSerAsnAsnThrSerThrAlaGlyAlaIleTyrVal 300
|||||
22249  TACTCTTCAGACAAATACATGACACAGCGGGAGGAGCTATCTATGTG 22298
301  LysLysLeuGlnLeuAlaSerGlyLysLeuThrLeuPheSerArgAsnSe 317
|||||
22299  AAAAAGCTCGAAGCTGGCTCCGGAGAGCTTACCTATTCAGTACAAATAG 22348
317  rValAsnGlyLysThrAlaProLysGlyAlaIleAlaIleGlnLys 334
|||||
22349  TGTCAAATGAGAGTACAGCTCTTAAAGGTGAGCCATACCTATCCAGATYA 22398
334  erGlyGluLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
|||||
22399  GTGGGGAATTAGTTATCCGCCATAGTGTGACATTGCTTTTATAGG 22448
351  AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
|||||
22449  AATACAGTCACTCTACTACTCTCGGAGCAATAGAACTAGTATGACACT 22498
367  uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleAlaLysArgAla 384
|||||
22499  AGGACAGAGTGCAAAGATGACAGCTTTCGCTGCTGCTGTAGAGCCA 22548
384  leTyrPheTyrAspProIleThrThrGlySerSerThrThrValThrAsp 400
|||||
22549  TCTACTTCTAGATCCCATACTACAGGATCATCCACACAGTTACAGAT 22598
401  ValLeuLysValAsnGluThrProAlaAspSerAlaLeuGlnTyrThrG 417
|||||
22599  GTCTTAAAGTTAATGAGACTCCGCGAGATTCTGCACACTCAATATACAG 22648
417  yAsnIleIlePheThrGlyLysLeuSerGluThrGluAlaIleAspS 434
|||||
22649  GAACATCATCTTCACAGAGAAAGTTATCAGACACAGGCCCCAGATT 22698
434  erLysAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGlyGly 450
|||||
22699  CTAAATAATCTTACTTCGAGAGTACTACGCTGTAACCTTTCAGAGGT 22748
451  ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
|||||
22749  ACTCTACTTTTAAACATGAGTACTCTGCAGACTCAGGCAATTCACATCA 22798
467  nGlnAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuGlnPro 484
|||||
22799  ACAGGCAAGATTCTGCTCTCGAATGAGAGCTAGCACTATCTAAGACCTG 22848
484  laAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
|||||
22849  CTGATTAACACCATTAACAAATTTGGTCAATTACATCAACATCTATAGAC 22898
501  GlyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThr 517
|||||
22899  GGTCCAAAGAACGCAAAATAGAACCAACCAAGCTACGTCAAAAATCTGAC 22948
517  rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGlu 534
|||||
22949  TTATCTGGAACATCACTTATATGAGCCGAGCGGAGCTTTATAGAAA 22998
534  snHisSerLeuAlaArgAsnProGlnSerTyrAspIleLeuGluLeuLysAla 550
|||||
22999  ATCATAGTTTAAGAAATCCAGAGCTTACGACATCTTAGAGCTCAAGCT 23048
551  SerGlyThrValThrSerThrAlaValThrProAspProIleMetGly 567
|||||
23049  TCTGAGACTGTAAACAACACCGCAGTACTCCAGATCTTAAATGGGTGA 23098
567  uLysPheHisTyrGlyTyrGlnGlyThrTyrGlyProIleValTyrGly 584
|||||
23099  GAAATTCATTACGGCTATCAGGGAACCTTGGGCCCAATTTGTTGGGGGA 23148
584  hrgAlaSerThrThrAlaThrPheAsnThrThrLysThrGlyTyrIle 600
|||||
23149  CAGGGCTTCTACAGCTCAACCTTCACTGAGCTTAACATGGCTATATT 23198
601  ProAsnProGluArgIleGlySerLeuValProAsnSerLeuThrAsnAl 617
|||||
23199  CCTATATCCAGAGGTATCGGCTCTTATAGTCCCTAATAGCTTATGGAATGC 23248
617  apheIleAspIleSerSerSerLeuAsnTyrLeuMetGluThrAlaAsnGlu 634
|||||
23249  ATTTATAGATTAATGCTCTCTCATTAATCTTATGAGAGCTCAAGCAAG 23298
634  yLeuGlnGlyAspArgAlaPheTyrPcysAlaGlyLeuSerAsnPhePhe 650
|||||
23299  GGTTCGAGGAGAGACCGCTTTTGGTGCTGCTGATTAATCTTCTTC 23348
651  HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGly 667
|||||
23349  CATAGAGATAGTACAAACACACAGCGGCTTCCGCAATTTGAGTGGCGG 23398
667  yTyrValIleGlyLysAsnLeuHisThrCysSerAspLysIleLeuSer 684
|||||
23399  TTATGCTATGAGAGAACTTACATCTTGTTCAGATTAAGATCTTAGTG 23448
684  laAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAlaLys 700
|||||
23449  CTGATTTTGTCAAGCTCTTGGAAAGATAGACACTTGTAGCTAAG 23498
701  AsnGlnGlyThrValTyrGlyGlyThrLeuTyrTyrGlnHisAsnGlu 717
|||||
23499  AATCAAGGTACAGTCAACGAGAGAACTCTTATACAGCACACAAACAAC 23548
717  rTyrIleSerLeuProCysLysLeuArgProCysSerLeuSerTyrValP 734
|||||
23549  CTATATCTCTTCTTCTTGCACAACTACGGCTTGTGTTTCTTATAGTTC 23598
734  roThrGluIleProValLeuPheSerGlyAsnLeuSerTyrThrHisThr 750
|||||
23599  CTACAGAGATTCTGTTCTCTTTCAGAGAACTTACGCTACACCACTACG 23648
751  AspAsnAspLeuLysThrLysTyrThrThrTyrProThrValLysGly 767
|||||
23649  GATPACGATCTGAAACCAAGTATACAACTATCTCTTATTAAGGAAG 23698
767  rTyrGlyAsnAspSerPheAlaLeuGlnPheGlyGlyArgAlaProIle 784
|||||
23699  CTGGGGAATAGTATGTTTCCCTTACAAATTCGTTGGAAGAGCTCCGATT 23748
784  yLeuAspGluSerAlaLeuPheGluGlnTyrMetProPheMetLysLeu 800
|||||
```

23749 GCTTAGATGAAGTCTCTATTGAGACAGTACATGCCCTTCATGAATTG 23798
 801 GlnPheValIYrAlaHisGlnGluGlyPheYssGlnGlnGlyThrGluAl 817
 |||||||
 23799 CAGTTTGTCTATGCATCAGGAAGTTTAAAGAACGAGGGAACGAAAGC 23848
 817 AARGGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGlyI 834
 |||||||
 23849 TCGTAATTTGGAAGTAGCCGCTCTGTGAATCTTCCCTTACTTACGGGA 23898
 834 leArgPheAspYssGlnSerAspYssGlnAspAlaThrThyAsnLeuThr 850
 |||||||
 23899 TCCGATTGATAGGAATCAGACTGCCAAGATGCACACTTACAACTTAAT 23948
 851 LeuGlyIYrThrValAspLeuValArgSerAsnProAspCysThrThrTh 867
 |||||||
 23949 CTGGTTACTGTGATCTTGTCTGTGTAACCCGACTGTAGACACAC 23998
 867 rLeuArgIleSerGlyAspSerTrpLysThrPheGlyIYrAsnLeuAla 884
 |||||||
 23999 ACTGGAATTAGCGGTGATCTTGGAAACTTTGCGTACGAATTTGGCAA 24048
 884 rGglnAlaLeuValLeuArgAlaGlyAsnHisPheCysPheAsnSerAsn 900
 |||||||
 24049 GACAACTTTTGTCTCTTCTGTCAGGGAACCAATTTTGTCTTAACCTCAAT 24098
 901 PheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSerArgAs 917
 |||||||
 24099 TTGAGAGCTTTAGCCAAATTTCTTTGAATGTGGTGGTATCTCCGAA 24148
 917 nTyraSnValAspLeuGlyAlaIYrGlnPhe 928
 |||||||
 24149 TTACAATGTAGACTTAGAGCAAAATACCAATTC 24182
 seq_name: gb_ba1:AB033795

seq_documentation block: 2241 bp DNA BCT 31-OCT-1999
 LOCUS AB033795 Chlamydomophila pneumoniae pmp-3.2 gene for polymorphic outer
 DEFINITION membrane protein G family, complete cds.
 ACCESSION AB033795
 VERSION AB033795.1 GI:6172316
 KEYWORDS polymorphic outer membrane protein G family.
 SOURCE Chlamydomophila pneumoniae (strain:U138) DNA.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
 REFERENCE 1 (sites)
 AUTHORS Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M.,
 Takeuchi,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H.,
 Matsushima,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A.,
 Ishii,K., Shiba,T., Hattori,M. and Kuhara,S.
 TITLE Comparison of outer membrane protein genes omp and pmp in the whole
 JOURNAL genome sequences of Chlamydia pneumoniae isolates from Japan and US
 REFERENCE Unpublished (1999)
 AUTHORS 2 (bases 1 to 2241)
 TITLE Hirakawa,H. and Shirai,M.
 JOURNAL Direct Submission
 REFERENCE Submitted (23-OCT-1999) to the DDBJ/EMBL/GenBank databases. Hideki
 AUTHORS Hirakawa, Kyushu University, Graduate School of Genetic Resources
 TITLE Technology: 6-10-1 Hakozaeki, Higashi-ku, Fukuoka, Fukuoka 812-8581,
 URL:http://www.grt.kyushu-u.ac.jp, Tel:81-92-642-3043,
 Fax:81-92-642-3043
 FEATURES
 source
 1. .2241
 /location/Qualifiers
 /organism="Chlamydomophila pneumoniae"
 /strain="U138"
 /db_xref="taxon:83558"
 /note="synonym:Chlamydia pneumoniae"
 1. .2241
 /gene="pmp-3.2"
 1. .2241
 /gene="pmp-3.2"

/note="containing frame shift"
 /codon_start=1
 /transl_table=11
 /product="polymorphic outer membrane protein G family"
 /protein_id="BA85948.1"
 /db_xref="gi:6172317"
 /translation="MSALFSENTSSKKGAQIOTSDAITLTGNGEVSFSDTSSDGA
 ALFTEASVTISNNKVSFIDNKRVTGASSSTGDSGGAICAKYKSTDTKYLNGNOML
 LEFNSSTTAGGAIVYKRLASGLTLFSRNSVNGGATPAKGAIAIEDSGELSLAD
 SGDIYELNVTSTTPGINRSSIDLGTSKWTALRSAGRAIYFDPITTGSSSTTV
 VLKVEITPADSALOYTGNIIFTEGKLSSTEAADSKNLTKLQPVTLGGTLSTKHG
 TLQOAFIQOADSRLNDVSTLEPADSTNTNVISSIDAKKAKIEFKASKNL
 TISGITITLDPGTGEYENHSIRNPOSYDILEKASGVTSVTAIYVPDPIMGKTHYGT
 GTWPIVWGTASTTATFNKIGTIRNPEITGSLVPSNLMNAIDISSLHYMETAN
 EGLQDRAFWCAGLSNFEHDKSTTRGRFRLISGVYIGNLIHTCSKILSAFCQLF
 GRBDYEPANKNQGVYGGTLYQHNETYISLPCRLRCSLSYVEIPIVFSGLSYT
 HTDNDLKRTYTPYTVKSGMNSDFALFEFGGRAPICDESALFEQYVPMKLDQVYH
 QEGFKEQGTAEAREFGSSRLVNLALPIGRDKESDQDQATYNIITGYTNLVSNPC
 TTTLRISDSKRTGTMARQALVLRAGNHFCSNFEAFSOFSELRGSSRNINDL
 GAKYQF"

BASE COUNT 660 a 474 c 478 g 629 t
 ORIGIN

alignment_scores:
 Quality: 3865.00 Length: 746
 Ratio: 5.181 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-428-122-2 x AB033795 ..
 Align seg 1/1 to: AB033795 from: 1 to: 2241

183 MetSerAlaLeuPheSerGluAsnThrSerLysGlyAlaI 199
 |||||||
 1 ATGTCAGCTCTGTTTCTGAAAAATACCTCCTCAAGAAAGCGGAGCAT 50
 199 eGlnThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerP 216
 |||||||
 51 TCAGACTTCGATGCCCTTACCATTAAGTGAACCAAGGGGAAGCTCTT 100
 216 hSerAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluAla 232
 |||||||
 101 TTCTGACAAATCTCTCTGATCTGAGCTGCAATTTTAAACAACACC 150
 233 SerValThrIleSerAsnAsnAlaIYrValSerPheIleAspAsnLysVa 249
 |||||||
 151 TCGGTGACTATTCTTAATAATGCTAAAGTTTCTTATTGACAAATAAGT 200
 249 lThrGlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaIleC 266
 |||||||
 201 CACAGAGGAGACTCTCTCAACACGCGGATATGTCAGAGAGTGTATCT 250
 266 yAlaIYrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGln 282
 |||||||
 251 GTCCTTATAAACTAGTACAGATTAAGTCAACCTCACTGCAATACAG 300
 283 MetLeuLeuPheSerAsnAsnThrSerThrAlaGlyAlaIleIYr 299
 |||||||
 301 ATGTTACTCTTCAGCAACAATACATGCAACAAGGGGAGGAGGTATCTA 350
 299 rValLysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerArg 316
 |||||||
 351 TGTGAAAAAGCTCGAACCTGCTTCGGAGGACTTACCTATTCAGTGA 400
 316 snSerValAsnGlyGlyThrAlaPpolysGlyGlyAlaIleAlaIleGlu 332
 |||||||
 401 ATAGTGCAATGAGAGTACACTCTTAAGGTGAGGACATAGCTATGAA 450
 333 AspSerGlyGluLeuSerLeuSerAlaAspSerGlyAspIleValPheLe 349
 |||||||
 451 GATAGTGGGGAATTGATTTCGCCGCAATGATGTGACATTGTCTTTT 500

349 uGlyAsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleA 366
 501 AGGGAATATACACTCTCTACTACTCTCGGACGAAATGAGATGATATG 550
 366 sPLeuGlyThrSerAlaIleMetThrAlaLeuArgSerAlaIleGlyArg 382
 551 ACTTAGGACACAGTGCANAGATGACAGCTTGCGTCTGCTGTGTGA 600
 383 AlaIleTyrrPheTyrrAspProIleThrThrGlySerSerThrThrValTh 399
 601 GCCATCTACTCTATGATCCATACACAGATCATCCACAACAGTTAC 650
 399 rAspValLeuLysValAsnGlyThrProAlaAspSerAlaLeuGlnTyrr 416
 651 AGATGCTTAAAGTTAATGAGACTCCGGCAGATTCCTGACTACAAATTA 700
 416 hrGlyAsnIleIlePheThrGlyGlyLysLeuSerGlyThrGluAlaIle 432
 701 CAGGGAACATATCTTCACAGAGAGAAAGTTATCAGACACAGAGCCGCA 750
 433 AspSerLysAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerG 449
 751 GATCTTAAAAATCTTACTTCGAGCTACTACAGCCTGTAACCTTTTCAG 800
 449 yGlyThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheT 466
 801 AGGTACTCTATCTTTAAACATGAGTACTCTCGACACTCAGGCAATTC 850
 466 hrGlnGlnAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuGlu 482
 851 CTCAACAGGCGAGATTCTGCTCGAAATGAGACGTAGAACTACTCTA 900
 483 ProAlaAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerI 499
 901 CCTGCTGATCTACACCATAAACAATTTGGTCACTTAACATCAGTTCTAT 950
 499 eAspGlyAlaLysLysAlaLysIleGlyThrLysAlaThrSerLysAsn 516
 951 AGACGTGCAGAAAGCAAAATAGAAACCAACCTACGTCMAAAATC 1000
 516 eutThrLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTy 532
 1001 TGACTTTATCTGGAACCATCTTATTTGACCCGACGCGCAGCTTTAT 1050
 533 GluAsnHisSerLeuArgAsnProGlnSerTyrrAspIleLeuGlnLeu 549
 1051 GAAATCATATAGTTAAGAAATCTCAGTCTACGACATCTTAGAGCTCA 1100
 549 sAlaSerGlyThrValThrSerThrAlaValThrProAspProIleMet 566
 1101 AGCTTCGGAACCTACACAGCAGCGCAGTACCCACATCTCTAATAG 1150
 566 LysGlyLysPheHisTyrrGlyThrGlnGlyThrProGlyProIleValTy 582
 1151 GTGAGAAATTCATTACGGCTATCAGGGAACCTGGGCCCAATGTTGG 1200
 583 GlyThrGlyAlaSerThrThrAlaThrPheAsnThrPheLysThrGlnTy 599
 1201 GGGACAGGGGCTTCTACGACTGCAACCTTCACAGGAGCTAAACGGCTA 1250
 599 rIleProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrp 616
 1251 TATTCCTAATCCGAGCGTATCGGCTTTAGTCCCTAATAGCTTATGGA 1300
 616 snAlaPheIleAspIleSerSerLeuHisTyrrLeuMetGluThrAlaAs 632
 1301 ATGCATTTATAGATATTAGCTCTCTCATTAATCTTAGAGAACGCAAC 1350
 633 GluGlyLeuGlnGlyAspArgAlaPheTrpCysLysGlyLeuSerAsnPh 649
 1351 GAAGGTTGCAGGAGACCGTCTTTTGGTGTCTGATATATCTAATCT 1400

649 ePheHisLysAspSerThrLysThrArgGlyPheArgHisLeuSerG 666
 1401 CTTCATAGAGATAGTACAAAACACACAGCCGGGTTTCGCATTTGAGTG 1450
 666 LysGlyThrValIleGlyLysLeuHisThrCysSerAspLysIleLeu 682
 1451 GCGGTTTGTTCATAGAGAGAACTACATCTTGTAGATAGATTTCTT 1500
 683 SerAlaAlaPheCysGlnLeuPheGlyArgAspArgAspTyrrPheVal 699
 1501 AGTCTCATTTTGTACGCTCTTGGAAAGATAGAGACTACTTGTAGC 1550
 699 aLysAsnGlnGlyThrValTyrrGlyThrLeuTyrrTyrrGlnHisAsn 716
 1551 TAAGAATCAAGTACAGTACGAGAGAACTCTCTATTACAGCAACAG 1600
 716 LuThrTyrrIleSerLeuProCysLysLeuArgProCysSerLeuSerTy 732
 1601 AAACCTATATCTCTCTCTCTGCAAACTACGGCCTTTGTGTGTCTAT 1650
 733 ValProThrGluIleProValLeuPheSerGlyAsnLeuSerTyrrTh 749
 1651 GTTCCTACAGAGATTCCTGTTCTCTTTCAGAAACCTTAGCTACACCA 1700
 749 sThrAspAsnAspLeuLysThrLysTyrrThrThrTyrrProThrValLys 766
 1701 TACGATACAGATCTGAAACCAAGTATACAACTATCTCTGTTAAG 1750
 766 LysErTrpGlyAsnAspSerPheAlaLeuGlnPheGlyGlyArgAlaPro 782
 1751 GAACCTGGGGAATGATAGATTGCTTACGATTCGTTGGAAGAGCTCCG 1800
 783 IleCysLeuAspGlnSerAlaLeuPheGlnGlnTyrrMetProPheMet 799
 1801 ATTTGCTTAATGAAAGTCTATTGAGACAGTACATGCTTCATGAA 1850
 799 sLeuGlnPheValTyrrAlaHisGlnGlnGlyPheLysGlyGlnTyrr 816
 1851 ATTGCACTTTGTCTATGACATCAGAAAGTTTAAAGAACAGGAGACAG 1900
 816 LuAlaArgGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIle 832
 1901 AAGCTCGTAATTTGGAAGTACCGCTCTTGTGAATCTTGCTTACCTAT 1950
 833 GlyIleArgPheAspLysGlnSerAspCysGlnAspAlaThrTyrrAsn 849
 1951 GGGATCCGATTTGATAGAAATCAGACTGCCAAGATCCACATCAATCT 2000
 849 uThrLeuGlyTyrrThrValAspLeuValArgSerAsnProAspCysThr 866
 2001 AACCTGTTGTTATCTGTGATCTGTTGTTAGTAACCCGACGTATGCA 2050
 866 hrThrLeuArgIleSerGlyAspSerTrpLysThrPheGlyThrAsnLeu 882
 2051 CAACACTGCAATTTAGCGGATTCCTTGGAACCTTCGTTAGCAATTTG 2100
 883 AlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisPheCysPheAsn 899
 2101 GCAGAGACACCTTAGTCTTCCTGCGAGGAGACCATTTTGTCTTAATCT 2150
 899 rAsnPheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSer 916
 2151 AAATTTGAGCCCTTTAGCCAAATTTCTTTGAAATTCGTTGGGCACTC 2200
 916 rGAsnTyrrAsnValAspLeuGlyAlaLysTyrrGlnPhe 928
 2201 GCATTTACAAATGTAGACTTAGAGCAAAATACCAATTC 2238
 seq_name: gb_pat: A81829
 seq_documentation_block: 2815 bp DNA
 LOCUS A81829
 DEFINITION Sequence 3 from Patent WO9585953.


```

1400 CCTGTATTATGATGCGGCACACGTTAAAGCAAGTACAGAGAG 1449
487 SerThiLeaAsnLeuValIleAsnIleSerIleAspGlyAla 503
1450 GTCACTTTAACAGGCTTTCCATTCTCTAGACTTTTAGCGGAGGTA 1499
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerg 520
1500 GAAAGTTGTAATGCTGCTTCGACAGCAAGTAAATGATAGCCCTTAGG 1549
520 LThrIleThrLeuAsnProThrIleGlyThrPheTyrGluAsnIleSer 536
1550 GTCCGATCTCTTTGGATACCAAGGAGACTTATGAATACACAGC 1599
537 LeuArgAsnProGlnSerTyrAspIleLeuGluLeuLysAlaSerGlyTh 553
1600 TTAGGAAAACACAGACTTTTCATTGTGCGACTCTGCTGCTGGGTAC 1649
553 rValThrSerThrAlaValThrProAspProIleMetGlyIulysPheH 570
1650 TGCAACACACTACAGATGTTCCAGCGGTTCTACAGTACCACTCCTACGC 1699
570 lSTYrGlyTyrGlnGlnGlyThrTPGlyProIleValTIP..... 582
1700 ACTATGGGTATCAAGTACTTGGGA...ATGACTTGGGTGATCATAC 1746
583 ..GlyThrGlyAlaSerThrThrAlaThrPheAsnTPPrThrLysThrG 598
1747 GCAACGACCTCCAAAGACTAAGACGACGACCTTACCTGACCAATACAG 1796
598 YThrIleProAsnProGluArgIleGlySerIleuValProAsnSerLeu 615
1797 CTACCTTCGAACTCTGAGCGTCAAGACCTTACCTTCCATAAGCTTT 1846
615 rPAsnAlaPheIleAspIleSerSerLeuHisTyrLeuMetGluThrAla 631
1847 GGGGACTTTTTCAGACATCCAGGATTCAGAGTGCATAGAGAGAGT 1896
632 AsnGluGlyLeuGlnGlyAspArgAlaPheTyrCysAlaGlyLeuSerAs 648
1897 GCTTGTGACTTTTGTTCAGATCGAGCTTCGCGCTGCGGAGTCCCAA 1946
648 nPhePheHisLysAspSerThrLysThrArgArgGlyPheArgHisLeuS 665
1947 TTTCTTAATATAAGTAAAGAAAGGGAAGCAAAATACCGTCAATAAT 1996
665 erGlyGlyTyrValIleGlyGlyAsnLeuHisThrCysSerAspLysIle 681
1997 CTGGTGATATGCTATCGAGGTCCAGCAAACTTGTGAAACCTTA 2046
682 LeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPheVa 698
2047 ATTACCTTTCCTTTTGGCAACTCTTGTGAGCGATAAAGATTTCTTAGT 2096
698 lAlaLysAsnGlnGlyThrValTyrGlyGlyThrLeuTyrTyrGlnHisA 715
2097 CGCTAAATATCATACTGATACCTATGACGAGCCCTCATATACCAACAC 2145
715 snGluThrTyrIleSerLeuProCysLysLeuArgProCysSerLeuSer 731
2146 .....ATTACAGAAATGATAGTGGGTTCATAGTGTCTCTTAAAT 2184
732 TyrValPro.....ThrGluIleProValLeuPheSerGlyAsnLe 745
2185 AAACCTCTGCTGCTTGGAGTCATAAACCCCTGTTTAGAAGGCGACCT 2234
745 userTyrThrHisThrAspAsnAspLeuLysThrLysTyrThrTrp 762
2235 CGCTATATAGCCACGTCAGTAAATGATCTGAAGCAATATATCTGCTATC 2284
762 rOrthValLysGlySerTyrPglLysAsnAspSerPheAlaLeuGluPheG 778

```

```

2285 CTGAGGTGAANGTCTTGGGGAATAATGCTTTAACTATGATGTGGGA 2334
779 GlyArgAlaProIleCysLeuAspGluSerAlaLeuPheGluGlnTyrMe 795
2335 GCTCTCTCTACTTCTTATCCGAATACCTGCATGTGTTTGTACTATGC 2384
795 tProPheMetLysLeuGlnPheValTyrAlaHisGlnGluGlyPheLysG 812
2385 TCCATATACATAACTGAATCTGACCTATATACGTCAGAGACGCTTCGG 2434
812 LuGlnGlyThrGluAlaArgGluPheGlySerSerArgLeuValAsnLeu 828
2435 AGAAGGTGACAGAGAGATCTTTGTATGACAGCAACCTCTTCAATTA 2484
829 AlaLeuProIleGlyIleArgPheAspLysGluSerAspCysGlnAspAl 845
2485 TCTTGGCTATAGGGGTGAAGTTGAGAACTTCCTGATTTATGACTT 2534
845 aThrTyrAsnLeuThrIleuGlyTyrThrValAspLeuValArgSerAsp 862
2535 TTTCTATGATCTGACTTATATCTATGTTCTGATCTTATCCGCAATGATC 2584
862 rAspCysThrThrThrLeuArgIleSerGlyAspSerTyrPheThrPhe 878
2585 CCAATGACATACAGCACTTGTATACAGCGAGCCTCTGGGAACCTTAT 2634
879 GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisP 895
2635 GCCAATAACTTACGACACAGACGCTTCAGTCCGTCAGCAGCACTGACTA 2684
895 eCysPheAsnSerAsnPheGluAlaPheSerGlnPheSerPheGluLeu 912
2685 CGCCTTCTCTCTGTTGTGAAGTCTGCGCAGTGTGCTTGAAGTTC 2734
912 rGlySerSerArgAsnTyrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
2735 GTGATCTCTACGAGATTATATGATGATCTTGGGGTAAGTCCAAATTC 2784

seq_name: gb_bai:AE002192

seq_documentation_block:
LOCUS AE002192 12676 bp DNA BCT 30-MAY-2000
DEFINITION Chlamydomophila pneumoniae AR39, section 25 of 94 of the complete
genome.
ACCESSION AE002192 AE002161
VERSION AE002192.2 GI:8163405
KEYWORDS
SOURCE
ORGANISM Chlamydomophila pneumoniae AR39.
REFERENCE 1 (bases 1 to 12676)
Bacteria: Chlamydiales; Chlamydiaceae; Chlamydomophila.
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
JOURNAL MEDLINE
20150255
REFERENCE 2 (bases 1 to 12676)
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7189226.
FEATURES
location/Qualifiers
1..12676
/organism="Chlamydomophila pneumoniae AR39"

```

gene
CDS
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(90. .185)
/gene="CP0300"
complement(90. .185)
/gene="CP0300"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38157.1"
/db_xref="GI:7189227"
/translation="MFTSLGRSAIFFTGSLKHLRGEFLFIQK"
complement(275. .1819)
/gene="CP0301"
complement(275. .1819)
/gene="CP0301"
/note="similar to GP:4376735; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF38158.1"
/db_xref="GI:7189228"
/translation="MTLRNFLTCSALFLPALPAQVYVYVHESDCYNGAIINNKSTLEPK
ITCYEGTGYIFLDVIRISNKHODEDAGVFIRSGNLFPMGRNCFTHNLMTEGFG
AISNRYGDTTLTLDSNFYSIAFTSAPILPDGGAGIYSLGSMVENSEEFTFCGNSSW
SGAATYPTLLSKASRSPVNSGNRIYVFRNVSOGYGAISTHNLTITTPGSCFE
NNHAYHDVNSNGGAIAPAGSISISVSGDIFLGNTASDQGNTHNSIHLQSAQF
KNRAVSESGVYADPISHESHKITDLYINAPEKEYTEGTSFSGCLDHEVCAL
NLITSLDQVTLAGGTLSDGVTLQLHSHFKQESSTLMSPTTLLSGDRAVNLH
LIIEPDNVPRIRAEADKSLAEKLVAFEAWSYVDPEQKEAFIPLLELQNP
SFDSLGLGETLERTQVTENDAVAGFMSLEVEYPPSLDKDRRTITPKKTYFLWNP
ETSTP"
complement(2077. .4926)
/gene="CP0302"
complement(2077. .4926)
/gene="CP0302"
/note="similar to GP:4376733; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF38159.1"
/db_xref="GI:7189229"
/translation="MIVLFCFYIDANSSLIKNSITMKTSIPVAVYSSVLAESCHQSL
ANELLSPDPSFNGNIDSGTTPKTSATITYSLTGDFYFEPEKGTPLSDSCFRQTTDN
LTFLENGSLITGFIIDAGTHAGAASTANKNLTFSGFLSFDSPSTVYVVGQTL
SSAGVNLNENIRKLVAGNFTADGGAIRKAGFLITGSGDALFENNSSSTKGAIAAT
TAGARIANNQGYRFLSNIASTSGAIDEGTISNNKFLYFEGNAAKTGACAT
KASGPELISNNKTLIFASNAVETSGGAIAHAKKIALSGGFTFLRANVSAPKNG
AISDASGELISAEQNTTPYRNLTITGSDTPKRNAININGSKGKTEFLPAAKNHT
IFPYDPTSEGTSSDYVLIKNNSAGALNPQGTIIFSGETTLADLKLAKADIKSSFTQ
PVLSGGKLLDQGVLTSTFSQAGSLGMDSTTLTSTAGSITITNLGIVDSILG
LKQVSLTPAKGASNKVIVSGKLNLIIDIGNIYESMFSHDQFLSLKITVADDTNV
DISLIPVPAEDPNSEYFQGMVNNMTDIAITNKATATWTGTFVSPERKCALV
CNTLGVFTDIRSLQQLVEIGALIGMEHKGFWNSMTNLHKTGDNKGRPRHSGGY
VITGSATPKDQDLTFEAFCHLFAKDQCFIANNNSRTYGTLFFKHSHTLOPQNLRL
GKAKSESAIEKFPREIPLADVOYFSHSDKRMETHTSLSPESGSNSECIAGIG
LDLPVLSNPHLPKFTFIIPQKVEVAVVYVSSFFSSSDGRGFSIGRLINLSIPGAK
FVQDIGDSYPLDLSGFVSDVYRNNPOSTAIIWVSPSWKIRGGLSRQAFLLRGSN
NVVYNSNCELFGHYAMELRGSSRNYNVDVGTCLRF"
5026. .7812
/gene="CP0303"
5026. .7812
/gene="CP0303"
/note="similar to GP:4376729; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"

gene
CDS
/protein_id="AAF38160.1"
/db_xref="GI:7189230"
/translation="MKSQFSLVYLSSTLACESTGTPAATNENIGSPDSFSTNRC
TTPKNTTGGIDYTLTGDTITLQNLGDSALIKGCFSDTTELSRKGYSISFLNIK
SAEGALSVTTDKNLISLTGFSLTFLAPSSVITTPSGKGAIVKCGDITFDNNGTILF
KQDYCEENNGAISTKNLSLNKSTGSIISPEGKSSATGKGAICATGTVDFNNATPT
LFSSNIAEAGAGAIINSTNCTITGNTSLVFESENSTYATAGAGALSGADVTISGNOS
VTFSSQAVANGAGIYAKKLLTLAGSGGGSIFSSNNIYOGTAGAGALSGADVTISGNOS
SAEAGDITFNGNAIYATTPORTKRSIDIGSTAKTINLRASISHSIPFYDITANTTA
DSTDITLNKADAGNSTDSSGSIYSSGKISSEDAKVDNLTSLKQVPTITAGNVL
KRGVTLDTKGTOTRAGSSVIMDAGITLAKSIEEYTLGLSIPVSLGKRVVIAASA
ASKVNLGSPILLDNOGNAYENHDGKTQDFSVQUSALDITATTTDPAVPATVAPT
HYVGQGTGMWVDOTASTPKTKATLAWNTGYLIPBERGQPLVPSLWMSFSDIQA
IOGVIERGALTLCSDPRGFAGVANEFLDKDKGKRYKRRHSGGAIAGSAQCSNL
ISPAFCQFGSDKDFLVAKNHTDPAAGFYLOHTEGSGFISGLDLKPGSMKPLV
LEGOLAYSHVNSDNLKTYTAYPEVKGSMWNAFMWIGASSHSXPEYILHCPDTPAYI
KLNTYIRQDSFSEGTGTRSFDDSNLENLSLIPGVKREKSDCNDPSYDITLSVPD
LIRNDPKCTIALVLSGASWETVANNLAQALQVRAHSHAFSPMFEVLGQFVEVRGS
SRIYNVDLGKRFQF"
complement(8170. .8292)
/gene="CP0304"
complement(8170. .8292)
/gene="CP0304"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38161.1"
/db_xref="GI:7189231"
/translation="WDLSSRWLNKLRKQDLQDNIKYCKSKYCLRGISQILGWV"
8329. .9444
/gene="CP0305"
8329. .9444
/gene="CP0305"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38162.1"
/db_xref="GI:7189232"
/translation="MEMWSPQPEQCHFDVYGSFLRPESITRARSDFEGRIVYGM
RVEDDAIRNLIRKQOTENAGLIFFTDGFERRYSMDPDMWGHGVDRRSDSDPEIGYV
LKDKISVSKHPIEFHEFYKTFEKNAGNAKQOTIPSPQFEMFIETAFNLKTRFYT
NOELIDIDIVEFYROVIOIDLYAACGRNLOLDCACRLIDRAPSVYVDSHDLROELL
EQLFIHINLWKKRDPEDLPFVSLHYCRDQYQAFPSRAYSISIEPLRAKDDVSYHXY
WALDDKISGSGKEPRLAYISGEKHVLCGLISSHSCIEDRDVAVSRITYEAASTIPLERLS
LSPQCGFASCGDHRMTEEBQWKIAFVKETAIKELWG"
complement(9785. .12571)
/gene="CP0306"
complement(9785. .12571)
/gene="CP0306"
/note="similar to GP:4376731; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF38163.1"
/db_xref="GI:7189233"
/translation="MKSLSLHFWLISLSLALPLSINFSAPAAVEINILGPTNSESQPT
YTPAQOTNMDAGTYLNLGQVSIINAGSPITALFASCKEETGNLSPGQHYQFLQNI
DAGNACFTNNTAAKRLSFGGSFVSLSIOTNATGCAIKSTACSIQSSQYFQO
NFSNDNGALOGSSISLSINDNLTFAKNKATOKGALYSGTITNNTLNASPENE
AANGGAIYTBASSFISNNKAISFINNSVATSTGAIYCSSISAKPVLITSDNE
LNFIGNRAITISGAIITYDNLVLSGGFTLRNNSAIDTAAPLDGGLAIADSGSLSLSA
LGSDITFEGNTVYKVGASSQITTRNSININTNNAKIVQLASQGNITFYDPIITTSIT
AALSDALNNGPDLAGNPAYOGTIVFSGERLSEAEADNLKSTIOOPTLLAGGQSL
LKSQVTLIVAKSFSSQSPSTILMDAGTILEADGITINLVANDSLKETKATLAKAO
ASQRTVLTSGLSITDPSGNYEDVSNPNPOVSLTLLADDPANIHITLDAADPLEKN
PIHWGYGNALNSQEDTATKSKAATITWTKTGYNPBERGTLVANTLWNGSPYDVAS
IQOLVATKROSQETKIGIWCIGISNFFHKDSTKINKCFRIISAGYVAGATTTLASDL
ITAAFCOLFGKDRDHFINKNRASAYASLHLQHLATLDRSSPLLRLYLGSSSESDPVLFD

AOISYISKNTMTYTTQAPKSSRYNDGCALELSPHLSHEGLFHAPEPK
 VEASVYHODSFKEKNTVLVSFDSGLINVSPIGTFEERSNERASVATVYVAD
 VYRKNPDCTALLINNTSMKTTGTNLRSROGIRAGIFAFSPNLEVTNLSMEINGS
 SRSYNDIGKPF"

BASE COUNT 3580 a 2439 c 2890 g 3767 t
 ORIGIN

alignment_scores: Length: 950
 Quality: 2058.00 Gaps: 20
 Ratio: 2.831
 Percent Similarity: 76.526 Percent Identity: 46.737

alignment_block:
 US-09-428-122-2 x AE02192 ..

Align seg 1/1 to: AE002192 from: 1 to: 12676

1 MetLysSerSerPheProLysPheValPhe...SerThrPheAlaIlePhe 16
 |||||.....
 5026 ATGAATCGCAATTTTCCTCGGTAGTCTCTCTGACATTTGGCATTTT 5075
 16 e.....ProLeuSerMetIleAlaThrGluThrValLeuAspSerS 30
 |.....
 5076 TACTAGTTGTTCCACTGTTTTCGTCGCAACTGCTGAAATATAGCCCT 5125
 30 eAlaSerPheAspLysAlaIleAlaIleGluAspSerValArgLys 45
 |||||.....
 5126 CTGATAGCTTTGACGAGTACTACACAGCCACTATCTCTCTAAAT 5175
 46 SerGluGluAspAlaGlyThrThrThrLeuPheLysGlyAsnValThrIle 62
 :|||.....
 5176 ACG.....ACACTGGAATAGACTATCTCTGACAGGAGATATACCT 5219
 62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsn 79
 :|||.....
 5220 GCAAAACCTT...GGGGAATTCGGCAGCTTTAACGAAAGGTTGTTTCTG 5266
 79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
 :|||.....
 5267 ACACACGCAATCTTAAAGCTTTCGGGTAGGGGTACTCCTCTCTTCT 5316
 96 GluThrValAspAlaGlyThrValAlaIleAlaIleValAsnSerSerVal 112
 :|||.....
 5317 TTAATATATAGTCT...AGTCTGAGGCGCAGCAGCTT...TCGTATAC 5360
 112 ValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleA 129
 :|||.....
 5361 AACTGATATAAATCTGCTGCTAACGAGATTTTCGAGCTTACTTCTTAG 5410
 129 LeuSerProGlySerSerIleThrThr.....GlyLysGlyValAlaSer 143
 :|||.....
 5411 CGGCCCTCATCGTATACACACCCCTCAGGAAAGGTGCACTTTAA 5460
 144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSe 160
 :|||.....
 5461 TGTGGA...GGGAGCTTACATTTGATTAACAATGCAATATTTATTAA 5507
 160 rLysAsnPheSerThrAspAsnGlyAlaIleAlaIleThrAlaLysThrLeuS 177
 :|||.....
 5508 ACAAGATTACTGTGAGAAATGGCGAGCCATTCTACCAAGAAATCTTT 5557
 177 eLeuThrGlyThrThrMetSerAlaLeuPheSerGlyAsnThrSerSer 193
 :|||.....
 5558 CTTTGAATAACGACGAGGATGATTTCTTTGAGGAGATATAATCGAGC 5607
 194LysLysGlyAlaIleAlaIleThrSerAspAlaLeuThr-11 207
 :|||.....
 5608 GCAACAGGGAATAAAGGTGGGCTATTGTGCTACTGCTGATGATGATAT 5657
 207 eThrGlyAsnGluGlyValSerPheSerAspAsnThrSerSerAspS 224
 :|||.....
 5658 TACAATATATAGCGCTCTACCTCTTCTCGAACATATATGCTGAGCTG 5707

224 eArgLysAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsnAla 240
 :|||.....
 5708 CAGGTGAGCTATATAATATGACACAGAACTGTACATATACAGGAAATPAC 5757
 241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerSerThrTh 257
 :|||.....
 5758 TCTCTGATTTTCTGAAATAGTGTACA.....GGACCCG 5795
 257 rGlyAspMetSerGlyGlyAlaIleCysAlaThrLysThrSerThrApr 274
 :|||.....
 5796 AGGA.....AATGAGAGGCTCT.....TCTGAGAGTG 5824
 274 hrLysValThrLeuThrGlyAsnGluMetLeuLeuPheSerAsnAsnThr 290
 :|||.....
 5825 CCGATGTTACCATATCTGGGAATCAGATGTAACCTTCTCAGAAACCAA 5874
 291 SerThrThrAlaGlyAlaIleThrValLysLysLeuGluLeuAlaSe 307
 :|||.....
 5875 GCTGAGCTAATGGCGAGCCATTATAGCTTAAGAGCTTACACTGCGCTC 5924
 307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyLys 322
 :|||.....
 5925 CGGGGGGGGGGGGTATCTCTTTCTTAACATATAGTCCACAGTACCA 5974
 322 hrAlaProLysGlyGlyAlaIleAlaIleGluAspSerGlyGlyLeuSer 338
 :|||.....
 5975 CTGACAGTAATAGTGGACCATTTCTATATACGCGAGCGGAGATGTAGT 6024
 339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValThrSe 355
 :|||.....
 6025 CTTTCAGCAGACGAGGAGACATTCCTCAATGGAATGCAATGCTATGTC 6074
 355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSerA 371
 :|||.....
 6075 AACACACACCAAACTCGAAAGAAATTCATATGACATAGACTACTAG 6124
 371 LeuLysMetThrAlaLeuArgSerAlaIleAlaIleThrPheThr 387
 :|||.....
 6125 CAAAGATACCAATTTACGTCATATCTGGCATAGCATATTTTCTTAC 6174
 388 AspProIleThrThrGlySerSerThrThrValThrAspValLeuLysVa 404
 :|||.....
 6175 GATCCGATTAATGCTAATATACGCTGCGGATCTACAGATTAATAATCT 6224
 404 LAsnGluThrProAlaAspSerAlaLeuGluThrThrGlyAsnIleIleP 421
 :|||.....
 6225 CAATTAAGCTGATGACGATATATACAGATTAATAGTGGGTGATGTT 6274
 421 heThrGlyGlyLysLeuSerGluThrGluAlaIleAlaAspSerLysAsnLeu 437
 :|||.....
 6275 TTCTCGTGAAGAACTCTCTGATAGATGACCAAAAGTTGCGAGAACCTC 6324
 438 ThrSerLysLeuLeuGluProValThrLeuSerGlyGlyThrLeuSerLe 454
 :|||.....
 6325 ACTTCTACGCTGCAAGCAGCCTGTAACCTTAACCTCAGGAAATTTAGTACT 6374
 454 uLysHisGlyValThrLeuGluThrGluAlaIlePheThrGluGluAlaAspS 471
 :|||.....
 6375 TAAAGGTGGTGTCTCTCGATACGAAAGGCTTACTACAGCCCGGCTT 6424
 471 eArgLeuGluMetAspValGlyThrThrLeuGluProAla...AspThr 486
 :|||.....
 6425 CCTCTGATATATGATGCGGCGACAAAGTAAAGCAAGACAGAGAGAG 6474
 487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLy 503
 :|||.....
 6475 GTCACTTTAAAGGCTTTCCATTCTGTACCTCTTAGGCGAGGCTAA 6524
 503 sLysAlaLysIleGluThrLysAlaIleThrSerLysAsnLeuThrLeuSerg 520
 :|||.....
 6525 GAAAGTTGTATATCTCTCTCTGACGACAGTAATAAATGTAGCCCTTAGTG 6574

NHMKTYTNDNSIIKSGSRMDACADGASLSPVIVSVYLLKREVEPVVVOYLYAHOOO
ETERRAEGRAFNKSELINYEIPICTFEEDSKSEKOTDILMTYIIDAIRRPKOOTIS
LTASDANMMAVGTNLAROGFSVRAAHFQVNPHEIFGFAEVSSSSKNYNTNLGSK
FCF"

sig_peptide 885..968
gene /gene="omp11"
CDS 885..3677
/gene="omp11"
/note="putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein 11"
/protein_id="CAB37068.1"
/db_xref="GI:4455883"
/translation="MKIPLKLLISSTLVTPILLISATYAGDASLPTDSFGAGST
FPRKSTADANGTNYVLSGNVYINDAGKGTALGCGCEFTERTGDTFTGKSPFMTVD
AGSNMAGSTADKALPFTGFSNLSFIAPGTTVASSEKSTLSAGALNLTMPGNTILF
SONVNEANNNGCATTTKTLISGNTSSTITFTSNKAKUGAIYSSAASISGNTGOL
VYNNKGTGGGALGEPASSITTONSLSFSGNTATDAGKGAICYCEKGTGPTLTI
SONKSLTFEENSSTVGGAIICAHGDLASAAGPTLFSNNCGMTAAGKGAIALIADSGS
LSLSANOGDITFIGNTLTSTSAPTSRNAIYLSSAKITNLKRAQOOSIYFYDPALSN
TIGASDVLTINQDPSNPLDYSGITVYFSGEKLADDEKAKADMTSLIKGPLASGTL
ALKNVELDVNGFTGEGSTLLMOPGTLKADTEALISLTKLVYDLSALEGKNSVSIET
AGAKNTITLSPVLFODSSGNFYESHITINQAFLOPIVFTATFAASDIYIDALLDIP
QTPPEHYGQGHMEATWADTSTAKSITMTVYTGQPNPERRASVYPSDLASFTDIR
TLOQIMTSONSITIOORGIMASGTANFERKDSGNQARHRKSYGIYVGSADSESEN
ITVAPCOLFGKDKDLFIVENTSHNLASLYIQHRAFLGGLMPSEFSITDMLKDIPL
ILNAOLSYSTYNDMDRTYTSYPEAOGSWTNSNGALGLOGSLALILPKAPFQCYFP
FLKFOAVYRQONFKESGAEARAFDGDLYNCSIPGAGILEKISEDEKNEFISLAYI
GDVYRRNPRTSLMWSGASMTSLICNLARQAFLASAGSHLTLSPVELSGAAVELR
GSAHYNDGGRYSP"

3827..6613
gene /gene="omp10"
CDS 3827..6613
/gene="omp10"
/note="putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="CAB37069.1"
/db_xref="GI:4455884"
/translation="MKSSLHMFLLISSSLALPLSNFSAFAVEINLGPNTSFGPGT
YTPAOTNADGTIYNLGDVGTINAGSPITALSCFKEGTGNSOOGYOFLQNT
DAGACOTNTNTANKLISFGFSYSLTIOTNTATGTGAIKSTGACSIOSNACFQNT
NFSNDGALGSSSISLSLNPULTRAKKATKGALYISGTCITINNTLNSAREBENT
NANNNGAITYEASSTISSKAISSFNNSTVATSAVGAIYCSSTSAKPVLLLSNGE
LNFIGNTALTSGAIYTDNLVLSGGPTLFKNNSAIDTAPLGAIAIDSGSLSLSA
LGDTFTFEGNTVVGKASSSOTTRNSINGNTAKIVOLRASOGNTIYFDPITTSIT
ALISDALNNGPDLNAGPAYOGTIVESGKLSAEAEAEADNLKSTIOOPLTAGOLS
LKSQYTLVAKSFSOSRGTLLMDAGCTLETTAGCTILNNLYLVNDSLKEKKTATLAKO
ASQYTLSSLSLVDPSGNVYEDVSNNNQVSCILTTDDPANIHITDLADPILKRN
PLHWIYQGNWALSWOEDATYKSKAATLTWTKGVNPNPERRGTVAANTLWGSFVRS
IOQLVATKYROSQETRGICWCEGISNEFHDSTKINKGFRHISAGVVGATTLADNL
ITAACOLGFKDRDHFINKNRASAYAAASHLQHLAKLISPSILRYLPGESSEOPVFLD
AOISYISKNTKTYTTOAPKGESSWYNDGALCLASSLPHTALSHEGLFHAYPEFI
VYASIIHDSPEKERTNTLIVRSFSDGLIVNSVPIGTEPERFERNERASYEAVIYVAD
VYRNKPDCTALLINNTSWKTGTCTNLSRQAGIGRAGIFTAFSPNLEVSNLSMETRGS
SRNIVNDGGRFQF"

3827..3907
sig_peptide /gene="omp10"
CDS complement(6954..8060)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB37070.1"
/db_xref="GI:4455885"
/translation="MSPFOQPEOCHFDVYGSFLRPESELTARSDPEEGRIYEQKRVY
EDAAIRNLIKOTEGALIFFTGDEFRARYSDMPWAGFCVDRRDSNDPELYGLKD
KTSVSKHPITEHFEFYKTEKNAKAKOTIPSPQEFHEMITAPNIKTRKRYPIPTNOE
LIDIVFYRYOVIODLYAAGCRNLQDDCAWCRLLDIRAPSWYGVDSHRLQELIEOF

LMIHNLWKKDREPDLFVSLHRCRGYQAEFFSRRARYDSIEERPLAKTIVDSHYHVAL
DKRISGAEPLAYSGERHVCGLGISNHSCEDEDDAVVASRIYENASTIPLERUSLSP
OCGFASCSGDHMTIEEOWKRIAFVKEIAKEIMG"

sig_peptide 815372
gene /gene="omp5"
CDS complement(8586..11372)
/gene="omp5"
/gene="omp5"
/codon_start=1
/transl_table=11
/product="outer membrane protein 5"
/protein_id="CAB37071.1"
/db_xref="GI:4455886"
/translation="MKSQFSLVLSSTIACFTSCGTVPATARENIGPDSFDGNTG
TTPKNTTGDYTLTGDTLQNLGDSALTRKGCSDTTELSLSEFKGKYSLSFLNKS
SAEGALSVTTDKNLSLTFGLFPLFSLPAPSVITTPSPCKGAVKCGGDLTPNNGLIF
KODYCEENGGAISTKNLSLKNSGSISEGNKSSATGKGAICATGVDITNNAPPT
LFSSNIAAAGAIINSTGNCITTGTSILVSESNSTVATAGNGALSGADVYISNOS
VTFSGNOAVANGAIYAKKLLILASGGGGISPSNNIYOGTTAGNGAISTILAGCSL
SEADGITFNCNAIYATTPOTTKRNSIDIGSTAKITTNRAISGHSIFPYDPTANTAL
DSTDLNLNKADAGNSTDYSGSIVFSGELSEDEKAVDNLTSLKQVTLTRAGLV
KRGVTLDRKGTOTRAGSSVINDAGTTLKASTEEVYTLGLSIPVSLGSKVYVLA
ASKNVALSGPTLLDNOONAAENHDLGKTODEFVOLSALOTATTDVAPVATVPT
HYGOGTGMWVDDTASTPRTKATFATLMTVYGLPNEROGPLVPSNLWCSFSDIOA
IOGVTERGALITLCSDRGFMAWAVANFELDKRKGKERYRHSGGYVIAIGAKOTCSEN
ISFARCOLFGSKDPLVAKNHDYTAAGFYIOHITGCGSGLITCLDKLPGSSMSHPLV
LEGOLYASHVNDLKTXYTAPEVYSGMGNNAFMNMGASSHSYPEYLHCDTAPYI
KLNTYIYKODSSEKGTGCRSFSDSNLNLSPITGVKEREKSDCHDSYDILTSTVPD
LIRNDPKTTLALVISGASWETVYANNLARQALQVRAGSHYAFSPMEVILGQFVFEVRS
SRNIVNDGGRFQF"

complement(11313..11372)
/gene="omp5"
11535..11588
/gene="omp4"
11535..14321
/gene="omp4"
11535..14321
/gene="omp4"
/codon_start=1
/transl_table=11
/product="outer membrane protein 4"
/protein_id="CAB37072.1"
/db_xref="GI:4455887"
/translation="MKTSIPWLVSSVLAFCSHLOSLANBELLSPDPSFGNDISGTF
TPKTSATYSLTGDFEYEPCKGPIPLSOGCKOTJDNLPFGNGSLSEFFGIDGTHA
GAASTGTGNTLTSFGSLSPDSSPSTTVTGLSDSAGVANNENTRIKLYVANGFS
TADGAIKAGAFLLITGSDGLFSSNNSSTGGATATACGARIANTYVFLSYIAS
TSGGAIIDEGSILSNKFLYFEGNAATGGAICONTAASGPELITSNKTLIPASN
VANTTGGAIHAKKILALSSGFTFERNVSSATPVGGAISIDASGELSLSAEVGTITF
VRNTLTGISTDTPKRNAINIGSNCKFTELPAAKNHTJFYDPIITSEGTSSDYKINN
GSAGLNLYOGCTILFSGETTLADELKVADNLKSSPTOPVSLGGKLLQKQVLTST
FSOEGSLNIGDSCGTLSTTAGSITITNLGINVDSGLKQVSLTAKGASMKVYSGK
LNLIDIBENIYESHMFSDQLFSLKITYVDADVPTNVDSILIPRAPDRPSEVFOG
QWNVWTTDITANIKETATVKTGFVSPERSKALVONTLMGFTDIRSLOQLVECHL
ATGHEWKQDVSWSNTNLRKTGDENRGRFTSGGYVIGSAHPKODLTDIRSLOQLVECHL
FARDSCDIANNNSRTYGTGLFPEKSHLQOQNYRLRQAKFESBALEKFPREIPLAL
DVQDFSDSNNMERTHYGLTLEFESGSMWNEIAGIIGLDLFPVLSNPPLKFTLPMQ
KVENYVYONSEFFESSSGRGESIGRLNLSIPGAKRVQGDIGDSYTYDLSGFEVSD
VYRNKPDCTALLINNTSWKTGTCTNLSRQAGIGRAGIFTAFSPNLEVSNLSMETRGS
SRNIVNDGGRFQF"

14579..14644
sig_peptide /gene="omp13"
gene 14579..16123
CDS /gene="omp13"
14579..16123
/gene="omp13"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB37073.1"
/db_xref="GI:4455888"
/translation="WTIRNFLTSCALFLAPAAQVYVYLHESDYGNAINKNSLEPK
ITCKPEGYSYFLIDVARSNVKHOEDAGVYINSNGIFPMGNCNCFNFHLLMDEGC
AALSIRVDDTTLTNSFSLAFTSNAPLLPQOGAIIYISGYSWIENSEVTVTCGNYSSE
SGAALTYPYLLGSKASRPSVNLSGNRVLYVFDVNSOGYGAISTHNLTLTRGPSCE

NNHAYHDVNSNGAIAIAPGGSISISVKSGLDFKNGTASODGNTIHNSIHLOGCAOF
 NLRVASESGVYFDPIESHSKHTDLVINAPESKEKEGHSISFSGLCDIDHEVCAR
 NLSTIIQDVLTAGSTLSLDVLTOLHSFKOEASSTLTMSGTTLCGDAVONLH
 ILIEDNDPEVPIRAEDKDALVSLERKVAFEAWSYSDPROFEARTIPLLELGP
 SPDSLLIGETTLERTQVTTENDAVGFMSLSNEERPSLDXDRITTPKTKTIFLTWNP
 EITSIP"

sig_peptide
 16494..16553
 /gene="omp14"
 16494..17280
 /gene="omp14"
 16494..>17280

CDS

alignment_scores:
 Quality: 2058.00 Length: 950
 Ratio: 2.831 Gaps: 20
 Percent Similarity: 76.526 Percent Identity: 46.737

alignment_block:

US-09-428-122-2 x CPN133034/rev ..

Align seg 1/1 to reverse of: CPN133034 from: 1 to: 17280

1 MetIySerSerPheProlySphenAlphe...SerThrPheAlaIlePh 16
 |||||
 11372 ATGAAATCGCAATTTCTGCTTCTCTCGACATTCGCACTTT 11323

16 e.....ProLeuSerMetIleAlaThrGluThrValIleuAspSers 30
 |
 11322 TACTAGTTGTTCCACTGTTTTCCTGCACTGCTGAATAATATAGCCCC 11273

30 eRLAserPheAspGlyAlaIleuAsn...GlyAsnPheSerValArgGlu 45
 |||||
 11272 CTGATAGCTTTGACGAGACTACACACAGCAGCTATCTCTTAAAAAT 11223

46 SerGluGluAspAlaGlyThrThrTyLeuPheGlyAsnValThrIle 62
 |||||
 11222 AGC.....ACTACTGGAATAGACTATCTCTGACAGAGATATTACTCT 11179

62 uGluAsnIleProGlyThrGlyThrAlaIleThrIySerCysPheAsnA 79
 |||||
 11178 GCMAAACCTT...GGGATTCGGCAGCTTTACGAAAGGTTGTTTCTG 11132

79 snThrIySGLAspLeuThrPheThrGlyAsnGlyAsnSerIleuLeuPhe 95
 |||||
 11131 ACCTACGCAATCTTAAAGCTTTGCGGTAAAGGGTACTCTCTTTT 11082

96 GluThrValAspAlaGlyThrValAlaIleAlaValAsnSerSeryA 112
 |||||
 11081 TTAATATTAAGCT...AGTCTGAAGGCGCAGCACTT...TCTGTTAC 11038

112 lValAspLySerThrThrPheIleGlyPheSerSerLeuSerPheIleA 129
 |||||
 11037 AACTGTAAAAAATCTGCTGCTACAGAGATTTGAGTCTTACTTCTTAG 10988

129 lAserProGlySerSerIleThrThr....GlyIySGLyAlaValSer 143
 |||||
 10987 CGGCCCATCATCGTAATCACAAACCCCTCAGAAAGGTGACAGTAA 10938

144 CysSerThrGlySerLeuSerLeuThrIyAsnValSerLeuLeuPheS 160
 |||||
 10937 TGTGGA...GGGATCTTACATTGATACAAATGCACTATTATTATA 10891

160 rLyAsnPheSerThrAspAsnGlyAlaIleThrAlaIySThrLeuS 177
 |||||
 10890 ACAGATTTACTGTGAGAAATGGCGAGCATTTCTACCAAGAACCTTT 10841

177 eRLeuthrGlyThrThrMetSerAlaLeuPheSerIleuAsnThrSerS 193
 |||||
 10840 CTTTGAAGAACAGCAGGAGTCTTTCTTGAAGCAATTAATGAGAC 10791

194LysIySGLyAlaIleGlnThrSerAspAlaLeuThrI 207
 |||||
 10790 GCAACGAGGAAAAAGGTGGGCTATTGTGCTACTGTACTGTATGATAT 10741

207 eRLySGLyAlaIleGlyIValSerPheSerAspAsnThrSerAspS 224
 |||||
 10740 TACAAATTAATACGGCTCTACCTCTTCTCGAACAATATTCGTGAAGCTG 10691

224 eRLyAlaIleAlaIlePheThrGluAlaSerValThrIleSerAsnAla 240
 |||||
 10690 CAGCTGAGCTATTAATACACAGAACTGTAACAATTACAGGAATATAG 10641

241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerSerThrTh 257
 |||||
 10640 TCTCTGTATTCTTGAATAATAGTGTGACA.....GCGACCCG 10603

257 rGlyAspMetSerGlyAlaIleGlyAlaIleCysAlaIyLysThrSerThAspT 274
 |||||
 10602 AGGA.....AATGAGAGACTCTT.....TCTGAGAGATG 10574

274 hRLyValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsnAsnThr 290
 |||||
 10573 CCGATGTACCAATATCTGGAATCAAGATGTAACCTTCTCAGAAACCA 10524

291 SerThrThrAlaGlyAlaIleTyValIySlyLeuGluLeuAlaSe 307
 |||||
 10523 GCTGATAGCTTAATGCGGAGCACTTATGCTAAGAACTTACACTGCTTC 10474

307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyLys 322
 |||||
 10473 CGGGGGGGGGGGGCGATCTCTTCTTCTACAAATATAGCCAGGATACA 10424

322 hRLAProlySGLyAlaIleAlaIleGluAspSerGlyLysLeuSer 338
 |||||
 10423 CTGCAAGTATATGTGAGCCATTCTTCTACTGCGACGAGAGAGTGTAGT 10374

339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValThrSe 355
 |||||
 10373 CTTTCAGCGAAGACAGGAGCACTTCACTTCAATGGCAATGCCATTGTTGC 10324

355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSera 371
 |||||
 10323 AACTACACACACAACACTCAAAAAGAAATCTATTGACATAGATGATCTAG 10274

371 lAlySerThrAlaLeuArgSerAlaAlaGlyArgAlaIleTyPheTy 387
 |||||
 10273 CAAGATTCACGATTATCGTGAATATCGGAGCATATCTTTTCTAC 10224

388 AspProIleThrThrGlySerSerThrThrValThrAspValLeuLysVa 404
 |||||
 10223 GATCCGATTACTGCTAATACGGCTCGGATTTCTACAGATCTTTAAATCT 10174

404 lAsnGluThrProAlaAspSerAlaLeuGlnTyThrGlyAsnIleIleP 421
 |||||
 10173 CAATAAGCTGATGACAGTAATAGTACAGATTATAGTGGTGTGATTGTT 10124

421 heThrGlyLysLysLeuSerGluThrGluAlaAlaAspSerIyAsnLeu 437
 |||||
 10123 TTTCTGGTGAAAGCTCTCTGAAATGAAGCAAAAGTTGCAACACCACTC 10074

438 ThrSerLysLeuLeuGlnProValThrLeuSerGlyLysThrLeuSerLe 454
 |||||
 10073 ACTTTACGCTTAACACAGCTGTAACTTACTACGACGAAATTTTGTACT 10024

454 uLySGLyAlaIleThrLeuGlnThrGluAlaPheThrGlnAlaAspS 471
 |||||
 10023 TAAAGTGTGTCTACCTCGATACGAAGAGCTTACTCAGACCCGGGTT 9974

471 eRLyLeuGlnIleuAspValGlyThrThrLeuGluProAla...AspThr 486
 |||||
 9973 CCTGTGTATTATGATGCGGCGACACAGTTTAAAGCAATACACAGAGG 9924

487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaIy 503
 |||||
 9923 GTCACTTTAACAGTCTTCCATTCTCTGTAAGCTTTTAAAGGAGGATTA 9874

seq_name: gb_bai:AP002546

seq_documentation_block:
LOCUS AP002546. 300600 bp DNA BCT 07-JUL-2000
DEFINITION Chlamydomophila pneumoniae genomic DNA, complete sequence, section 2/4

ACCESSION AP002546 AB033782 AB033783 AB033784 AB033785 AB033800 AB033801
AB033802 AB033803 AB033804 AB033805 AB033807 AB033808
AB033809 AB033810 AB033811 AB033812 AB033814 AB033815
AB036079 AB036080 AB036081 AB036082 AB038348 AB038349 BA000008
AP002546.1 GI:8978640

VERSION
KEYWORDS
SOURCE
ORGANISM Chlamydomophila pneumoniae (strain:J138) DNA.
Chlamydomophila pneumoniae
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE
1 (sites)
AUTHORS Shiba,T., Hirakawa,H., Kimoto,M., Tsuchi,M., Kishi,F., Ouchi,K.,
Shiba,T., Ishii,K., Hattori,M., Kohara,S. and Nakazawa,T.
TITLE Comparison of whole genome sequences of chlamydia pneumoniae J138
from Japan and CML029 from USA
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)
MEDLINE 20330349
REFERENCE 2 (bases 1 to 300600)
AUTHORS Shirai,M.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.
Mutsunori Shirai, Yamaguchi University School of Medicine,
Department of Microbiology, 1-1-1 Minamiyogushi, Ube, Yamaguchi
755-8505, Japan (E-mail:ms Shirai@pc.cc.yamaguchi-u.ac.jp,
Tel.:81-836-22-2227, Fax:81-836-22-2415)
COMMENT AB033782-AB033785, AB033800-AB033815: Submitted (25-Oct-1999)
AB038348-AB038349: Submitted (14-Feb-2000)
AB036079-AB036082: Submitted (18-Dec-2000).

FEATURES
SOURCE location/Qualifiers
1. 300600
/organism="Chlamydomophila pneumoniae"
/strain="J138"
/db_xref="taxon:83558"
/note="synonym:Chlamydia pneumoniae"
74. 865
/gene="CpJ0267"
74.. 865

CDS

```
/gene="CpJ0267"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA98477.1"
/db_xref="GI:8978641"
/translation="MSMSLNKTNALLNPPEPAVCLNANDPKYINODRKTFACVYTL
VATLMIITGVIVLLANGSPGLSVSTIIIGTSVLTGLTALFIITGLVLIKKSLAMI
OYKGOEYVQKYEKPEFSI PKNDNVKHLISCLPSPLDIESPSEASTPYSKRIACSG
VAIVGLVTLIGAVSVFCTGYLQIALVCGACIACLTALFVCGLAGLRHSLIAQGM
YIYLYIYSSALREERNEIVKDOENIYITLIECRQKREKALLE"
887..1273
/gene="CpJ0268"
/gene="tdk"
887..1273
/gene="CpJ0268"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BA98478.1"
/db_xref="GI:8978642"
/translation="MSOCSSSTSTWEMKSFVPMKNPTPLSPISPEDEFLIAYEP
FVLPTDPENAOANPGTSTPVGNGIDOLNLLQOPNQNNANNPGTSGSNPTSLPA
PRLPETEENSQEEQSGONNEDLIG"
complement(1425..2402)
/gene="CpJ0269"
/complement(1425..2402)
/gene="CpJ0269"
/complement(1425..2402)
/codon_start=1
/transl_table=11
/product="dipeptidase"
/protein_id="BA98479.1"
/db_xref="GI:8978643"
/translation="MTIDMHCDSLSPHFCRKDPARVRCSPQELLSGVROQCAIFV
HSRGPNCDKONSLEFSLPNQYDGLISYEENSGSSOKSLISRSIENASLGD
DIAPGLTLAKLIHLTKOGLAYLGIWKGDMRFGCGTEAPKRLSNDCKVLLDITVEL
GYPLDISHCSDKLADIDYTDKTPINLAVISHNFPSVDLHRNLDAHAKELVLR
KQVILNLVRSVVGSLDGLDEKHLVAEHLGLSSIVGSGDFYANEDENEFNECCSS
AEAPVLNQLHRIFSKKAESILSSRAEKFLOKQIVBOVNPKITDVL"
complement(2417..3277)
/gene="ywlC"
/complement(2417..3277)
/codon_start=1
/transl_table=11
/product="SuA5 superfamily-related protein"
/protein_id="BA98480.1"
/db_xref="GI:8978644"
/translation="MPDKAOITFSLPEVMSAIHQKIVALPTDVVGFVSLIASEA
EERLYALDKREPSKAFALVNSIEDIENISGPIPLTKKLAQLFPFGAITLVKHNRP
RPKPTTLARIVDHSVREIVDHGCTLIGTSANISEPFSALTAEITFADFDHDLCIF
DGPCSHLESTVVASDPLIYREGIISRSVTENAGTEAKTFHRTSHAFSKHIXITY
KNQDLVSFLSGSDFKGVCEHPRKMFYRLREALKKTPSIVFIYDINTSYPEL
FPFLSPYIE"
3586..4317
/gene="CpJ0271"
3586..4317
/gene="CpJ0271"
/codon_start=1
/transl_table=11
/product="lysophospholipase esterase"
/protein_id="BA98481.1"
/db_xref="GI:8978645"
/translation="WTDYSFFRRKIGNIEALECPGNPDPIILCHGYGSLADNLTFEF
PSICSFKLRPTWTFNGILPLENDFRGSACFPNLVLLQELSRLVANGVGNLOEKY
DELFDVLETKALEELIILNRPYNEIIIGFSOGAILATHLVLSQNYAGALIEF
AGARLFGNGWEGEIKQCOAVPFLQSHGEDLILPHLGAHLNDLLIKLNQPFSEHG
GHEIJSVYFOCKWQTVPMWIDPARG"
complement(4289..5185)
/gene="dnaX_2"
/complement(4289..5185)
/gene="dnaX_2"
/codon_start=1
/transl_table=11
```

```
/product="DNA polymerase III gamma and tau"
/protein_id="BA98482.1"
/db_xref="GI:8978646"
/translation="MHEEENQGWAEALRKRYHOEVPAILLHGFTLPVLQDKAEOLA
SEILLSPPGSESHVSOXIHPIDIVQFEFGGRHSIDIPGIKKOIYISPEFANXI
YIIEADRMTLAAISAFKVEEPPKHAVIILTTAKORLEKTIISGSLSTFIERGK
ILGSEKTEPFRYAOCQIPVTEYSOIKESSEFQDKOYLRKVRPMDVLELIRDX
TLNGLKASALNYEPRKHEIILQLEPLPLDKVLLIVESCSGLNNSSSAAVLEWVAIO
LVSIOYKEKELVSPGODUSN"
complement(5176..5796)
/gene="tdk"
complement(5176..5796)
/gene="tdk"
/codon_start=1
/transl_table=11
/product="thymidylate kinase"
/protein_id="BA98483.1"
/db_xref="GI:8978647"
/translation="MFIYIEGEGSGKSSLAKALGDLQVADRKVILLTREPQGLICE
RLRDLLEPPLHELRCCELPFLGSRAOHTOPVILBALDGYIVTICERHDSITVYO
GIAGLGDAPFADCSKVYGPFPPLPPLVLLDIPADIGLQRKHQKVFDFEKKPLS
YHNIRGEFLSLASADSRILVLDRESLASLIDKVMHLTQLGCT"
complement(5801..8305)
/gene="gyrA_1"
complement(5801..8305)
/gene="gyrA_1"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit A"
/protein_id="BA98484.1"
/db_xref="GI:8978648"
/translation="MFNKDETIYVKNLLEEMKESLYRSMSVITSRALPIDRGLKPS
ORRLVYMKOISLSPGAKHRCARIGTCDGDVPHESVYPIVYVADNNWARYLT
VDGQVYKMSIDSGIPAPARKTERALTISMTILMEDLDKDVYDVIPNDYERKHEPVP
SKFENLCONSGSGLAVGMATNIPPHNIGELIEATLULLAPQASVDEIIQVWGPQEP
TGIIICGSEGIIRSTYTTGRGKIKVRAHLEENDEKRESIIITEMYNNKSLLEQ
IANNVNETKLAGISDVDESDKDGIIVALEIKGSESEIILNRYKFTDVQVFGAM
LADKNLPRMISIRHMSAMIRHREKVIYRRRTRELRKATRAHVLEGYIKALSCDA
LVKTIRESGKHEAKERLIESFGTEQALAIILRLYOLTGLEAKIOXEVELINK
IAYKOYLSDEGLVKDIIREMLDILKHKHVARRTIEFPADIDREDIITNDSYIT
TISGDDYKRPKPVKVFKEQRGGGVTFVKKAGGLKAVYSAFTKDYLLITNFGQ
CYWLKWQLEGEGERAKRPIINLEGIIRGEELAAILNKNFDMAGFLFLATRGVY
KVSYLDAFSNPRKGIKALIDEGDELIACHIVISDEKVMFLTHLGMAYRPFHEVYR
PMGRTAGVGRVSLKNEDKEDVSCQIVTENOSVLYVCDGFGKSLVEDERETNRCGV
GVRSLIINENGVNLGAIPVTDHDSIILMSOGQAIRINNQDVVRWGRSTQGVRLVHL
KEGDALYSMEKLSNENDEVLGSGSEECSDTYSLR"
complement(8321..10738)
/gene="gyrB_1"
complement(8321..10738)
/gene="gyrB_1"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="BA98485.1"
/db_xref="GI:8978649"
/translation="MDPEKNYDASAITVLEGLQAVREPRGYTGDGTIGLHNLVEE
LVNNSIDEMAGYCSRIDVRILEDGGIVYIDNRGPIIEVHERESAKOGREYALVY
LTVLHAGKFPDSDSKYSGVLHGVSVAUNSEKIVAVAFVKCKYKQOMFSGSIVPT
PLQYVYSDRQGEIIVYDPKTIESTQCTEATLRLRELATLNGITITVPEDDDV
SEPKVTFYFGGIGOSFYSYIINQKESIFSEPIITICTRGDDSEIFEALQWNSGTA
ELVYSTANNIPTRQGGHILGFSYALTRVINYIKAHNLKANLKLAGEDEIREGLTA
VISEVKNPQFEQITQKLGNSDVSSAAQVWGEALTIFFEENPQIARIMVDFVFAA
QAREAAKARELILRKSLDASARLPGLKIDLCDEKDEKEMYIVEDSGAGSAGKQGRD
RRQVAILPIGKILNVEAKLOKIFPNOEIGTIIAAGGIGADNENLSKLYRRRII
MTQADVDSGILRIILLTFPRHNTALINDCVYIAQPLKYVKKKKDFYITSEKMD
SYILMIGTNSSLTFKSTERELGELSESTINVIDVESINTLEKKAIPSEFLEMY
KEGIGPLVYLAATGMOGGRVLYSDEEKEEALAOETKFKIIELYKAAVFEDIONQ
LKEIGYGLDISSYLIPOKNEIYVIGNEDSPSCVYLTLEVINYLNKNGRKGIEIQRKG
LGEMNADQMDWTMNEORTLIIHVSJKDAVEADHIPTMLGEEVPRPREFIESHALSI
RINNLDT"
complement(10742..11089)
/gene="CpJ0276"
complement(10742..11089)
```

```
/gene="CPj0276"
/codon_start=1
/transl_table=11
/product="Ct191 hypothetical protein"
/protein_id="BAA98486.1"
/db_xref="GI:8978650"
/translation="MFLKRRKGGSGVONKRTASPIKAKHYLHLYLOELQIMARP
HVAIDAMNOVFEDKTRKMSQAGFDHILLVKNYSYALLKQFPONDLMISLYOVA
SHVOIREIQFLG"
complement(11353..11862)
/gene="CPj0277"
complement(11353..11862)
/gene="CPj0277"
/codon_start=1
/transl_table=11

alignment_scores:
  Quality: 2058.00      Length: 950
  Ratio: 2.831          Gaps: 20
  Percent Similarity: 76.526  Percent Identity: 46.737

alignment_block:
  us-09-428-122-2 x AP002546/rev ..

Align seg 1/1 to reverse of: AP002546 from: 1 to: 300600

1 MetlySerSerPheProlysPheValPhe...SerThrPheAlaIlePh 16
  |||||.....|
207760 ATGAAATCGCAATTTCTCGTTAGTCTCTCTCTCTGACATGCGATGTT 207711
16 e.....ProleuserMetIleAlaThrGluThrValLeuSpsers 30
  |||||.....|
207710 TACAGTTGTTCCACTGTTTCTGCTGCACTGCTGAAATATAGCCCT 207661
30 eAlaSerPheaspGlyAsnLysAsn...GlyAsnPheSerValArgLys 45
  |||||.....|
207660 CTGATACCTTTCAGCAAGTACTTAACACAGCAGCTTACTCTAATAAAT 207611
46 SerClnclunspAlaGlyThrThrTyrLeuPheLysGlyAsnValThrLe 62
  ::|||.....|
207610 ACG.....ACTACTGGAATAGACTACTCTGACAGAGATATAACTCT 207567
62 uGluAsnIleProGlyThrThrAlaIleThrLysSerCysPheAsn 79
  |||||.....|
207566 GCAAAACCTT...GGGATTCGGCAGCCTTACACAGAGCTTTCTTCG 207520
79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
  ::|||.....|
207519 ACACCTACGGAATCTTTAAGCTTTCGCGTAAAGGGTACTCATTCTTT 207470
96 GlnThrValaspAlaGlyThrValAlaGlyAlaAlaValAsnSerSerVa 112
  ::|||.....|
207469 TTAATATTTAAGTCT...AGTGTGAAGGCGCACACTT...TCTGTTAC 207426
112 lValaspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleA 129
  ::|||.....|
207425 AACTGATATAAATCTGTCCGTACACAGATTTTCAGTCTACTTCTTAG 207376
129 lAsnProGlySerSerIleThrThr.....GlyLysGlyAlaValSer 143
  |||||.....|
207375 CGGCCCATCATATCGGTATACCAACCCCTCAGCAAAAGGCGAGTTAAA 207326
144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSe 160
  |||||.....|
207325 TGTGCA...GGGATCTTACTTTGATACACAAATGCAACTATTATTATTA 207279
160 rLysAsnPheSerThrAspAsnGlyAlaIleThrAlaLysThrLeus 177
  ::|||.....|
207278 ACAAGATTACTGTGAGGAAATGCGGAGCCATTCTTACCAAGATCTTT 207229
177 eTrleuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer 193
  |||||.....|
207228 CTTTGAAGAAACGACGCGATCTTTCTTTGAAGGAATATAATCGAGC 207179
```

```
194 .....LysLysGlyAlaIleGlnThrSerAspAlaLeuThrIle 207
  |||||.....|
207178 GCAACAGCAAGAAAAAGCTGGGCTATTGTGCTAGTGTACTAGTAT 207129
207 eThrGlyAsnGlnGlyLysValSerPheSerAspAsnThrSerSerAspS 224
  |||||.....|
207128 TACAAATAATACGCGCTCTACCTCTCTGCAACAAATTTGTGAAGCTG 207079
224 eGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsnAla 240
  ::|||.....|
207078 CAGGTGACGTATTAATATAGCACAGAACTGTACATTAAGGGAATAGC 207029
241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerThrThr 257
  ::|||.....|
207028 TCTCTGTATTCTTGAAATAGTGTGACA.....GCGACCCG 206991
257 rGlyAspMetSerGlyAlaIleCysAlaIleThrLysThrSerThrAsp 274
  ::|||.....|
206990 AGGA.....AATGGAGAGCTCTT.....TCTGGAGATG 206962
274 hrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsnAsnThr 290
  ::|||.....|
206961 CCGATGTACCATATATCGGGAATCAGAGTACTTCTCAGGAAACCA 206912
291 SerThrThrAlaGlyAlaIleThrValLysLysLeuGluLeuAlaSe 307
  ::|||.....|
206911 GCTGTACTAATGCGCGAGCCATTATGCTAAGAACCTTACACTGCTTC 206862
307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyLys 322
  |||||.....|
206861 CGGGGGGGGGGGGGATATCTCTCTTCTTACATATAGTCCAAAGTACCA 206812
322 hrAlaProLysGlyGlyAlaIleAlaIleGluAspSerGlyLysLeuSer 338
  |||||.....|
206811 CTCACAGTATATGTTGGGCGCATTTCTATACCTGCACCTGAGAGTGTAG 206762
339 LeuSerLysPheSerGlyAspIleValPheLeuGlyAsnThrValThrSe 355
  |||||.....|
206761 CTTTCACAGACAGACAGGACATTTACCTTCAATGGAATGCCATTTGTC 206712
355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSerA 371
  |||||.....|
206711 AACTACACCCCAACAACTACGAAAGAAATTTCTATTTGACATGATCTAC 206662
371 lAlysMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleThrPheThr 387
  |||||.....|
206661 CAAGATCACAGAAATTTACGTCGAATATCTGGGATGACATCTTTCTTAC 206612
388 AspProIleThrThrGlySerSerThrThrValThrAspValLeuLysVa 404
  |||||.....|
206611 GATCCGATTTCTCTAATACGCTGCGGATTTCTACGATATCTTTAAATCT 206562
404 lAsnGluThrProAlaAspSerAlaLeuGlnThrGlyAsnIleLeP 421
  ::|||.....|
206561 CAATTAAGCGTGATGACAGGATATGATACATTTATAGGGTGCATTTGTT 206512
421 hethrGlyGlyLysLeuSerGlyThrGluAlaAlaAspSerLysAsnLeu 437
  |||||.....|
206511 TTTCTGTGTAAGGCTCTGGAAGATGAACAAAGTTGGAGCAACCTC 206462
438 ThrSerLysLeuGlnProValThrLeuSerGlyGlyThrLeuSerLe 454
  |||||.....|
206461 ACTTCAAGCTGTAAGCAGCCTGTACTCTACACGAGGAATTTATATAC 206412
454 uLysAsnGlyValThrLeuGlnThrGlnAlaPheThrGlnAlaAspS 471
  |||||.....|
206411 TAAACGGGGTGTCTCTGATCGAAAGGCTTTACTCGACCCGCGGT 206362
471 eArgLeuGlnMetAspValGlyThrThrLeuGlnProAla...AspThr 486
  |||||.....|
206361 CTTCTGTATTATATGATGCGGCGACAAACGTTAAACCAAGTACAGAGAG 206312
```


YTPAOTNADGTYVNLGVDVSTNAGSPATLASCFKENTGNLSFQCHGYOFLONI
DAGACTPTNTANKLISFSGFSLIOTTNATTTGAKTSGACISQNSICFQ
NFSNDNGALQSSISLSNPLTFKAKNATQGGALYSTGGTTINTNLSASFENT
AANNNGAIYTEASFSISNKAISFINSSTGATGALYCSSTPKVLLTSDVS
LNEIGNTAITSGAIYTDNLSSGPTLFKNNAIDPAPLAGALAIADSGLSA
LGDDIFEGNTVVGKASSQOTTRNSINGNTAKIVOLRASGNTIYVDPPTTSIT
AASDLNNGPDLAGNPACVOCITVSGEKLSEAEADNLSKSTIOQPLTAGGOLS
LKSQVTLVAKSFQSPGSTILMDAGTTLLETADGTTINLVANDSKETKATLAKIQ
ASQVTLSSLSLVDPGSVYEDVSWNNQVESCULITADDPANIHITDLADPLEKN
PIHMGYQNMWALSMDIATKSKAATLTWTKGYNPNPERGTLVANTITLSDVRS
IOLVATKYROSETRGICEGISNFKHDSKINKGPHISAGYVAGTTLASNL
ITAFQGLFGKDRDHFINKNRASVAAASHLOHLATLSPSLIRYLPSSSEDEPVL
AOLSYKTKMTKTYTOAPKGSWYNDGCALIASPHTLSHSGEFLHATFPIK
VEASTIHQDSFKERTTLVRSFSDGLINVSPTGITTFERSHNERASTEATVIVAD
VIRKNPDCITALLINFTSKTGTGTLNSROAGIGRAGIFYAFSPNLEVTNLSMEIRGS
SRSYNADLGKRFQF

gene
/gene="yxjg_2"
/note="CpH0448"
complement(3233..4348)
/gene="yxjg_2"
complement(3233..4348)
/gene="yxjg_2"
/codon_start=1
/transl_table=11
/product="Hypothetical Protein"
/protein_id="A018592.1"
/db_xref="GI:4376732"

CDS
/translation="MEMSPFOOPROCHFDVSGFLRPESTLRASDFEGRIVROM
RYVEDAIRNLIKKOTEGALIFTDGEFRYSDEDFMGFQVDRRSDNDEIVY
LNDKISVSHNPIEHEFVKTEFKNAKAKOTIPSSQFHEHIEFNPKNRKPEPT
NOELIDDIIVYRQVIOQLYAAQRNLOLDCMCRILDI RASWYSGSHRLOEIL
EOLFIHNLVMPKRPEDLEVLHRCGDYQAEFSRRADVISLEPFAFDSDSYRY
WALDKYSGAEPRLAYSGEKLHVCGLISNHSICIDPRAVYSRIEASYSIPLEBS
LSPOGCFASCEGDRHMTFEEMOKKIAFVKEIKKEING"

gene
/gene="pmp_10.1"
complement(4865..6769)
/note="CpH0449"
complement(4865..7650)
/note="Polymorphic Outer Membrane Protein G Family;
contains frame shift"
complement(6715..7650)
/gene="pmp_10.2"
/note="CpH0450"
/gene="pmp_10.1"
813..10599
/note="CpH0451"
813..10599
/gene="pmp_11"
/codon_start=1
/transl_table=11
/product="Polymorphic Outer Membrane Protein G Family"
/protein_id="A018593.1"
/db_xref="GI:4376733"

misc-feature
/translation="KRTSIPWLVYSVLAFSCHLOSILANELLSPDSENGINDSCT
TRTSATYSLTGDVFEYRPGKPLSDSCFKOTDNLFTLGLSHLTGTFIDAGHA
GAAASTTANKNLTFSGSLSDSPSTVTGOGNLSAGVNLNINIKLVAGNS
TADGCAIKASFLTFSGDALFNSNSTGKALITVAGARIANTGVRPLNSAS
TSGCAIDEGTSLSNKFLYEGNAGKTTGACIKYKASGSELIISNNKILIFNS
VETSGCAIHAKKLASGTFEFLNNVNSAPKGCASIDASGSLISAPGNTF
VANTLTGSGTDTPKRNAINIGSKFTFLRAKNTITFTFPIETEGISDYLKIN
GSGALNPYOGTILFSGETLVADLDELVADNLASFQPVLSGGKLLDOKVLEBS
FSQEAQSLIGMDSGTLSTTAGSITITNGIIVDSIGLQVSLTAKGASNVIVSG
LNLIDEGNIVESHMFSDOLFSLKITVDADVDNVDISLIPVAEDPNSGFG
QNNVNTTDTATNTKATATWTKTGFVSPERSALVCNTLNGVFDISLQOLVIG
ATGMEHKEGFVWSMTNPLHKTGDNKRGFRHSGGVYIGSAGHTPKDLFARFOL
FARDKOCFIAHNNSRTYGTIFPKHSHTOPONXYLIGAKSSESIEKFPHEIPAL
DVQVFSHSDNRMETHTYSLPESBGSWNECAGIGLDPFLVLSHPLFTFTIPOM
KVMYVYQNSFFESSSDGRGFSIGRLNLSIPVCAKFVQGDIGSYTYDLSGFEVSD
VYRNPNQSTATLVMSDPDSWKIRGNLSROAFLRGSNNVYNSNELFCHYAMEIRGS
SRNYNDVGTIKLRF"

BASE COUNT 3228 a 2459 c 2065 g 3005 t
ORIGIN

alignment_scores:
Quality: 2044.00 Length: 950
Ratio: 2.804 Gaps: 20
Percent Similarity: 76.737 Percent Identity: 46.737

alignment block:
us-09-428-122-2 x AE001628/rev ..

Align seg 1/1 to reverse of: AE001628 from: 1 to: 10757

1 MetlysSerPheSerPolySphaValPhe...SerThrPheAlaIlePh 16
|||||.....
7650 ATGAATCGCAATTTCTCTGTTAGTCTCTCTCGCATTTGGCATGTT 7601

16 e.....ProLeuSerMetIleAlaIleThrGluThrValLeuAspSer 30
TACTAGTTGTTCCACGTGTTTCTGTCACACGTCGAAATATAGCCCT 7551

30 eAlaSerPheAspGlyAsnLysAsn...GlyAsnPheSerValArgGlu 45
|||||.....
7500 CTGATACCTTTGACGGAAGTACTACACAGCAGCTTACTCTTAAAT 7501

46 SerGlnGluAspAlaGlyThrThrTyrLeuPheLysGlyAsnValIle 62
ACG.....ACTACTGGAATAGACTATACCTGACAGAGATATACTCT 7457

62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsn 79
|||||.....
7456 GCAAAACCTT...GGGATTCGCGACCTTAACCAAGGTTGTTTCTG 7410

79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
|||||.....
7409 ACATACGGAATCTTTAAGCTTTCGCGTAAAGGCTACTCATTTCTTT 7360

96 GlnThrValAspAlaGlyThrValAlaGlyAlaValAsnSerSerVa 112
|||||.....
7359 TTAATATTAAGTCT...AGTGTGAGCGGACGACCTT...TCTGTTAA 7316

112 lValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIle 129
|||||.....
7315 AACGATAAATCTGTCGCTAACAGGATTTTCAGCTTACTTTCTTAG 7266

129 lAserProGlySerSerIleThrThr.....GlyLysGlyAlaValSer 143
|||||.....
7265 CGGCCCATCATCGTATCATCAACCCCTCAGAAAGATGCGATTAA 7216

144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPhe 160
TGTGCA...GGGATCTTACATTTGATACATGCAATGCAATTTTATTTAA 7169

160 rLysAsnPheSerThrAspAsnGlyValAlaIleThrAlaLysThrLeu 177
|||||.....
7168 ACAAGATTACGTGAGGAAATGCGGACATTTCTACCAAGATTTT 7119

177 eLeuThrGlyThrThrMetSerAlaLeuPheSerGlyAsnThrSerSer 193
|||||.....
7118 CTTGAAACACAGCAGGATCGATTTCTTTGAAGGAAATTAATCAGC 7069

194LysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrI 207
|||||.....
7068 GCAACAGGAAAGAAAGTGGGCTATTGCTACCTGCTAGATAT 7019

207 eThrGlyAsnGlnGlyValSerPheSerAspAsnThrSerSerAsp 224
|||||.....
7018 TACAAATAATACGGCTCTCTCTCTCTGACAAATATGCTGAACCTG 6969

224 eArgLysAlaIlePheThrGluAlaSerValThrIleSerAsnAsnAla 240
|||||.....
6968 CAGGTGACCTATTAATATGACAGAAACGTACAAATTAACAGGAATACG 6919

241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerThrTh 257
|||||.....
6918 TCTCTGTATTCTCGAAATAGTGTGCA.....GCGACCGC 6881

553 tValaIserSerThrAlaValThrProAspProIleMetIyluLysPheH 570
 6002 TCGCAACAACTACAGATGTTCCAGCGGTTCTTACATATACCACTCCCTACGC 59535
 570 1stYrGlyGlyGlnGlyThrTrpGlyProIleValTrp..... 582
 5952 ACTATGGGTATCAAGTACTCTGGGA...ATGACTGGGTTGATGATACG 59068
 583 ...GlyThrGlyAlaSerThrThrAlaThrPheAsnTrpThrLysThrG1 598
 5905 GCAAGCACATCCAAAGACTAAGACACGACGACATTAAGCTTGACCAAAATACAG 58566
 598 YTYrIleProAsnProGluArgIleGlySerLeuValProAsnSerLeuTr 615
 5855 CTACCTTCGGATCCTGAGCGTCAGAGACGCTTTACTCTTAATACCTT 58068
 615 rPasnAlaPheIleAspIleSerSerLeuHisTYrLeuMetGluThrAla 631
 5805 GGGGATCTTTTTCACACACATCCAAAGCATTCAGAGTGCATAGAGAGAGAACT 57566
 632 AsnGluGlyLeuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAs 648
 5755 GCTTGACCTCTTGTCTGCATGAGAGGCTTTGGGCTGGGAGATCGCCA 57068
 648 nPhePheHisTyrAspSerThrLysThrArgArgGlyPheArgHisLeuS 665
 5705 TTTCTTAGATTAAGATTAAGAAAGGGGAAAAAGCGAAATATCCGTATAAAT 56566
 665 eTCGlyGlyTYrTrpAlIleGlyGlyAsnLeuHisThrCysSerAspLysIle 681
 5655 CTGGTGGAATGTCATCGAAGGTGACGCGCAAACTGTTCGAAAACTTA 56068
 682 LeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAspTYrPheVa 698
 5605 ATTAGCTTGGCTTTGGCAACTCTTGGTAGCGATTAAGAATTCCTTAGT 55566
 698 lAlaLysAsnGlnGlyThrValTYrGlyGlyThrLeuTYrGlnHisA 715
 5555 CGCTAAATATCATCTGATGACTTACCTTAGCGAGGCGCTTCATATCCAAAC 55070
 715 nGluThrTYrLeuSerLeuProCysLysLeuArgProCysSerLeuSer 731
 5506ATTACAAATGTAGTGGGTTCCATAGTGTCTCTTAGAT 54688
 732 TYrValPro.....ThrGluIleProValLeuPheSerGlyAsnLe 745
 5467 AAACCTTCCTGGCTCTGGAGCATTAACCCCTCGTTTAGAAGGCGAGCT 54188
 745 uSerTYrHisThrAspAsnAspLeuLysThrLysTYrThrTrp 762
 5417 CGCTATATAGCCACGTCATATGATCTGAAACAAAGTATATCTCGATTC 53688
 762 rOThrValLysGlySerTrpPheLysAsnAspSerPheAlaLeuGluPheGly 778
 5367 CTGAGGTGAAGAGTCTTGGGGAAATATGCTTTAATCATGATGTGGGA 53188
 779 GlyArgAlaProIleCysLeuAspGluSerAlaLeuPheGluGlnTYrMe 795
 5317 GCTTCTTCATTTCTTATCCGAATACCTGCATTTGTTTGTATACCTATGC 52688
 795 rProPheMetLysLeuGlnPheValTYrAlaHisGlnGluGlyPheLysG 812
 5267 TCCATACATCAAACTGAAATCTGCACCTATATACGTCAGAGACAGCTTCTCGG 52188
 812 lGlnGlyThrGluAlaArgIubPheGlySerSerArgLeuValAsnLeu 828
 5217 AGAAAGGTACAGAGGAAGATCTTTGTGTCACACAACTCTTCATTTTA 51688
 829 AlaLeuProIleGlyIleArgPheAspLysGluSerProCysGlnAspAl 845
 5167 TCTTTGGCTATAGGGGTGAATTTTGAAGAAGTCTCTGATTTGATGACTT 51188
 845 aThrTYrAsnLeuThrPheGlyTYrThrValAspLeuValArgSerAsn 862

```
.....
5117 TTCTTAGATGACTTATTCCTATGCTCTGATCTTATCCGAATGATG 5068
862 roAspCySthrThrThrLeuArgLleSerGlyAspSerTrpLysThrPhe 878
5067 CCNAATGACATACAGCACTTGATATCAGCGAGCCTTTGGAAACTAT 5018
879 GlyThrAsnLeuAlaArgLlnAlaLeuValLeuArgAlaGlyAsnHisPh 895
5017 GCCAATTAACCTAGCAGCAGCAGCCTTCAAGTGCCTCAGCAGCATCTA 4968
895 eCySpheAsnSerAsnPhcGlnAlaPheSerGlnPheSerPheGlnLeuA 912
4967 CGCCTTCTCTCCTATGTTTGAAGTGGCTGGCACTTGTCTTGAACTTC 4918
912 rGlySerSerArgAsnTrpAsnValAspLeuGlyAlaLysTrpGlnPhe 928
4917 GTGATCTCAGCGATTTATATGATAGATCTTGGGGTAAAGTTCATTC 4868
seq_name: gp_pat:A81837

seq_documentation_block:
LOCUS A81837 2757 bp DNA PAT 21-JAN-2000
DEFINITION Sequence 11 from Patent WO9858953.
ACCESSION A81837
VERSION A81837.1 GI:6731869
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 2757)
AUTHORS Madsen, A. and Birkelund, S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL Patent: WO 9858953-A 30-DEC-1998;
MAIDEN ANNA SOFIE (DK); BIRKELUND SVEND (DK)
FEATURES
source 1..2757
Location/Qualifiers
/db_xref="taxon:332644"
BASE COUNT 770 a 564 c 642 g 781 t
ORIGIN

alignment_scores:
Quality: 2001.00 length: 953
Ratio: 2.814 Gaps: 21
Percent Similarity: 74.607 Percent Identity: 45.435

alignment_block:
US-09-428-122-2 x A81837 ..
Align seg 1/1 to: A81837 from: 1 to: 2757

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
1 ATGAGATCGCTTTTCTTCTTATTAATCTTCACTCTAGCCCTTTC 50
17 cLeuSerMetIle.....AlaThrGlnThrValLeuAspSer 30
51 TCTCTTAATGAGTCTTCTGCAATGCTGCCGATCTCACAATTAGGGAGTC 100
30 eTrLaserPheAspGlyAsn...LysAsnGlyAsnPheserValArgGlu 45
101 GTGACAGTTTATGATGATACAAAGCCACAGATTACTCTCTAAAGCG 150
46 SerGlnGlnAspAla...GlyThrThrThrLeuPheLysGlyAsnValTh 61
151 GCAACTTCTGATGCTAGTGCAGACCTATATCTCATGAGGATGCTC 200
61 rLeuGlnAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheA 78
201 GATTAAGCAA...GCAGGGAACAACAGACCTTAACCAAGTTGTTT 247
```

```
78 sAsnThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeu 94
248 CTACACACTCAGCAAAATCTTACTCTTAGGCAACGATTTCTCTTCAAT 297
95 PheGlnThrValAspAlaGlyThrValAlaGlyAlaValAsnSerSe 111
298 TTGACAAATATTTTTCGTACTGTTGCAAGGTGTTGTTGTTGACATAC 347
111 rValValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPhe 128
348 ACCAGCTCTGGGATTAACGAATCTTCAGAGATTTCACACTCTCGGATC 397
128 LeuLaserProGlySerSerIleThrThrGlyLysGlyAlaValSerCy 144
398 TTGACACTCTTAGG.....ACCACAGTAAAGAGCCATTAATAAT 438
145 SerThrGlySerLeu.....SerLeuThrLysAsnValSe 156
439 ACCGATGCTCTGTGTTGAGAGTATAGGAATCTTGACCAAAAT.... 483
156 rLeuLeuPheSerLysAsnPheserThrAspAsnGlyAlaIleThrA 173
484 .....GAAATGCTCTAGTGAATAATGGGAGGACCATCAATA 520
173 lAlaSerThrLeuSerLeuThrGlyThrThrMetSerAlaLeuPheSerGlu 189
521 CGAAGACTTGTCTTGTGACTGGAGTACCGGCTTGTAGCTTCTTGGC 570
190 AsnThrSerSerLysGlyGlyAlaIleGlnThrSerAspAlaLeuTh 206
571 AATAGCTGTCGCAACAGAGGAGCATCTATGCTTGGTGGACTCTGT 620
206 rIleThrGlyAsnGlnGlyGluValSerPheSerAspAsnThrSerA 223
621 GATTTGTGAATGACAGCAATCTTGACTTCGAAACAACACATGCGACAA 670
223 sPserGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsn 239
671 CATCAGAGAGCGGATCTCTGCGAAGGACACTGTGATCTCCATTAAC 720
240 AlaLysValSerPheIleAspAsnLysValThrGlyAlaSerSerTh 256
721 CAATAATATCTTTTCATGATGCTGCAACAGCACT..... 753
256 rThrGlyAspMetSerGlyGlyAlaIleCysAlaTrpLys.....Thrs 271
754 .....ACAAATGGCGAGCATTTGATTGATTAACAAGACAGGGCGA 793
271 eThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSer 287
794 ACCCAACCCCTATCTTGACTCTTTCAGAAATGAGAGCCTGCATTTTC 843
288 AsnAsnThrSerThrAlaGlyAlaIleIleTrpValLysLysLeuGl 304
844 AATAACACGACGAGAAATAGTGGAGGTGCGATTATACCAAAAATTGCT 893
304 uLeuAlaSerGly...GlyLeuThrLeuPheSerArgAsnSerValAsnG 320
894 GTTATCTCAGAGCGAGGAGAGTGTATTATTCTAACAACAAGAGCTGCA 943
320 lGlyThrAlaProLysGlyAlaIleAlaIleGlnAspSerGlyGlu 336
944 ATGCTACT...CCTAAGAGAGGGCAATTGCGATTCTAATCTTGAGAG 990
337 LeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThyA 353
991 ATTAGATTTCGACAGATCTCGCAATATCATTTTCGAGGGAATAACT.. 1038
353 lThrSerThrThr.....ProGlyThrAsnArgSerSerIleAspL 367
1039 .ACGACACTACAGGAAGTCTGGAGGTGACACGAATGCTATAGATC 1087
367 euGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleGlyArgAla 383
```

```

1088 TTGCATCGAATGCAAAATTTTAATCTCCGAGGAGCTCGGGGAAATAA 1137
384 ILeryrherlyrapprollethThrGlySerSerThrValThrAs 400
1138 GTTATTTTCTATGATCCTATCAGC.....AGCTCAGAGAGCTACTGA 1178
400 pValLeuLyValAsnGlnThrProAlaAspSerAlaLeuGlnThrG 417
1179 TAGCTCTCTTGTGATAAGCTGACGAGATCTGGAATACTATGAA 1228
417 LysnIleIlePheThrGlyLysLeuSerGlnThrGlnAlaAsp 433
1229 GCTACATCGTTTCTCTGAGAGAACTCTCAGAGAGAACTTAAGAA 1278
434 SerLysnLeuThrSerLysLeuGlnProValThrLeuSerGly 450
1279 CCTGACATGTGAATCTATACAGCTTACAGCTTGTCTGACG 1328
450 YThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
1329 TGCCTTAGTATGTAAGATGAGTGAAGTACTGCAATACATAAGCC 1378
467 InGlnAlaAspSerArgLeuGluMetCaspValGlyThrThrLeuGlu... 482
1379 AGGTGAGGAGTGAAGTGTATGATGAGGAGGAGTACTTGTAGGCA 1428
483 ProAlaAspThrSerThrIleAsnLeuValIleAsnIleSerSerI 499
1429 AGCGCTGAGGGGGTCACTCTCAATGGCTTACCATTAATATAGATTCTT 1478
499 eAspGlyAlaLysLysAlaLysIleGlnThrLysAlaThrSerLysAsn 516
1479 AGATGGGACAAATAAAGCTATCATTAAGGCGCAGCAGCAAGTAAGAG 1528
516 euthrLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTy 532
1529 TTGCTTATCAGGGGCTATCATGCTTGTAGATGCTCAGGGGAATTTAT 1578
533 GluAsnHisSerLeuArgAsnProGlnSerTyArgIleLeuGlnLeu 549
1579 GAGCATCATTAATCTCAGTCACAGCAGGCTTCTTATATAGAGCTTTC 1628
549 sAlaSerGlyThrValThrSerThrAlaValThrProAspProIleMe 566
1629 TCCACAGAGACGATGATCTACATACAGATATCCCGATACCCCAATCTTA 1678
566 LysGluLysPheHisTyArgLysGlnGlyThrThrProLysProIleVal 582
1679 ATACTACGAACTATGGGATCAAGAACT...GGAATTAATTTGTTGG 1725
583 .....GlyThrGlyAlaSerThrThrAlaThrPheAsnThrPheTy 596
1726 GTCGAGATGCAACTGCAAAACAAATAATGCTACTTAAGCTTA 1775
596 sThrGlyTyThrProAsnProGlnArgIleGlySerLeuValProAsn 613
1776 AACGAGATACAAAGCCGAATCCAGAACTGACGAGCTTGTGTTCTTA 1825
613 eLeuThrPheAsnAlaPheIleAspIleSerSerLeuHisTyLeuMetG 629
1826 GCGTGGGGGCTTCTTTCGATGTCGCTCATATCAGAGCTCATAGGAC 1875
630 ThrAlaAsnGlnGlyLeuGlnLysAspArgAlaPheThrProCysAla 646
1876 CGGACACAAAGTTGTTACTTCTGTCACAAATTTGTGGTATCAGGA 1925
646 uSerAsnPhePheHisLysAspSerThrLysThrArgArgGlyPheArg 663
1926 CGCGGACTTTTTCATGAGATCAGAAAGAAACCAACGTAATTTTCGTC 1975
663 lSLeuSerGlyTyThrValIleGlyLysLeuHisThrCysSerAsp 679

```

```

1976 ATTCTAGCCCGGCTATGCCATTAGAGAGGATTTCTTACGGCTTCTGAA 2025
680 LysIleLeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAsp 696
2026 AATTCCTTAATTTGCTTTTGTGAGCTTTTGGCTACGACAAAGACCA 2075
696 rPheValAlaLysAsnGlnGlyThrValTyArgLysThrLeuTyArg 713
2076 TCTTGGTGAAGAACCATCCATGATATATCCAGGGGCAATGAGCTTAC 2125
713 InHis.....AsnGlnThrTyIleSerLeuProCysLysLeuArg 726
2126 GACACCTCGAGAGTCTAGACCTCCGTAAGATT..... 2160
727 ProCysSerLeuSerTyValProThrGlnIleProValLeuPheSerG 743
2161 .....TTGTCAGAGAAATCTGATCCCTTACTTGTGCTTCATAGC 2201
743 LysnLeuSerTyThrHisThrAspAsnLeuLysThrLysTyThrT 760
2202 TCGGTTTGCTATGGCCATACCGACAATACATGACCAAGATACACTG 2251
760 hTyThrProThrValLysGlySerTyArgLysAsnAspSerPheAlaLeu 776
2252 GCTATTCCTCCTGTTAAGGGAAGCTGGGAATGATGCTTCGATAGAA 2301
777 PheGlyGlyArgAlaProIleCysLeuAsp...GluSerAlaLeuPheG 792
2302 TGTGAGAGAGCTATCCCGTATGCTTTCAGAGAGCTCCGCTTGGGTGA 2351
792 uGlnTyMetProPheMetLysLeuGlnPheValTyThrAlaHisGln 809
2352 TACCCACAGCCCATTTCAACCTAGATGATCATATCAGATAGAAATG 2401
809 LysPheLysGlnGlnGlyThrGlnAlaArgGlnPheGlySerSerArgLeu 825
2402 ACTTAAAGAAACGACAGAGAGCGCTTCTTCCAAGAGGAAGACTTC 2451
826 ValAsnLeuAlaLeuProIleGlyThrLeuPheAspLysGluSerAsp 842
2452 TTCAATCTAGCGGTTCTGTAGGATTAATTAAGAAATTTCTCCGAT.. 2499
842 sGlnAspAlaThrTyAsnLeuThrLeuGlyTyThrValAspLeuValA 859
2500 ....AAGCTACGATATGATCTCTCCATAGCTTACGTTCCGATGTGATC 2545
859 rGSerAsnProAspCysThrThrThrLeuArgIleSerGlyAspSerTy 875
2546 GTAATGATCCAGGCTGCACGACAACTTTATGATGTTCTGGGGATTTCT 2595
876 LysThrPheGlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAla 892
2596 TCGACATGTGTACAGGCTTGTCTAGACAAAGCTCTTGTGACGTGCTGG 2645
892 LysnHisPheCysPheAsnSerAsnPheGlnAlaPheSerGlnPheSer 909
2646 AAATCATCATCTCTTGTCTCAAACTTTGAAGTTTCAAGCAGTTTGAAG 2695
909 heGlnLeuArgGlySerSerArgAsnTyArgAsnValAspLeuGlyAla 925
2696 TCGAGTTGCGAGGTTCTTCTGATGCTATGCTATGCAATTTGGAGGA 2745
926 TyArgInPhe 928
2746 TTGCGATTT 2754

```

```

seq_name: gb_pat:A81851
seq_documentation_block: 3000 bp DNA
LOCUS A81851
DEFINITION Sequence 25 from Patent WO9858953.
ACCESSION A81851
VERSION A81851.1 GI:6731876

```



```

|||||.....:|||||.....:
1408 GATCGATTACTGCTAATACGGCTCGGATTCCTACAGATACCTTAAATCT 1457
404 1asnGluThrProAlaAspSerAlaLeuGlnTyrGlyThrGlyAsnIleLeu 421
|||||.....:|||||.....:|||||.....:
1458 CAATAGAGCTGATGCAGATTAATAGTACAGATTATAGTGGTCGATGTTT 1507
421 heThrGlyGluLeuSerGluThrGluAlaAlaAspSerLysAsnLeu 437
|||||.....:|||||.....:|||||.....:
1508 TTTCGGTGAAGAGCTCTCGAATGAAAGCAAAAGTGCAGACACCTC 1557
438 ThrSerLysLeuLeuGlnProValThrLeuSerGlyGlyThrLeuSerLe 454
|||||.....:|||||.....:|||||.....:
1558 ACTTCTACGCTGAAGACAGCTGTACTCTTACTCAGAGAAATTAGTACT 1607
454 uLysHisGlyValThrLeuGlnThrGlnAlaPheThrGlnGlnAlaAsp 471
|||||.....:|||||.....:|||||.....:
1608 TAAACGTGGTGTCACTCGATACGAAAGGCTTTACTCAGACCCGGGTT 1657
471 eArGLeuGluMetAspValGlyThrThrLeuGluProAla...AspThr 486
|||||.....:|||||.....:|||||.....:
1658 CCTGTATTATATGATGCGGGCACACGTTAAAGCAAGTACAGAGAG 1707
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAla 503
|||||.....:|||||.....:|||||.....:
1708 GTCACTTAAACAGGCTTTCCATTCCTGTAGACTCTTTAGCGCAGGATA 1757
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerG 520
|||||.....:|||||.....:|||||.....:
1758 GAAAGTGTATATGCTGCTCTCTGTCAGCAGATAAATAGTACCCCTTAG 1807
520 LyrThrIleThrLeuLeuAspProThrGlyThrPheTyrGluAsnHisSer 536
|||||.....:|||||.....:|||||.....:
1808 GTCCGATCTCTTTTGTGATACCAAGGAATGTTATGAAAATTCACGAC 1857
537 LeuArGAsnProGlnSerTyrAspIleLeuGluLeuLysAlaSerGlyTh 553
|||||.....:|||||.....:|||||.....:
1858 TTGAGGAAATCAAGACTTTTCATTTGTCAGACTCTCTCTCTGGGTAC 1907
553 rValThrSerThrAlaValThrProAspProIleMetGlyGluLysPheH 570
|||||.....:|||||.....:|||||.....:
1908 TGCAAACAACTACAGATGTTCCACGCGTTCCTACAGTACGAACTCTAC 1957
570 1stYrGlyTyrGlnGlyThrTrpGlyProIleValTrp.....582
|||||.....:|||||.....:|||||.....:
1958 ACATGGGTATCAAGTACTGGGGA...ATGACTTGGGTGATGATACC 2004
583 ...GlyThrGlyAlaSerThrThrAlaThrPheAsnTrpThrLysThrG 598
|||||.....:|||||.....:|||||.....:
2005 GCAAGCACTCCAAAGACTAAGACAGCAGCATTTAGGCCAATACAG 2054
598 yTrpIleProAsnProGluArgIleGlySerLeuValProAsnSerLeuT 615
|||||.....:|||||.....:|||||.....:
2055 CTACCTTCGGAATCTGAGCGTCAAGACCTTTAGTTCCTAATAGCCCTT 2104
615 rPAsnAlaPheIleAspIleSerSerLeuHisTyrLeuMetGluThrAla 631
|||||.....:|||||.....:|||||.....:
2105 GGGGATCTTTTCAGACATCCAGCATTCAGAGTCTCATAGAGAAAGT 2154
632 AsnGluGlyLeuGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAs 648
|||||.....:|||||.....:|||||.....:
2155 GCTTTGACCTTTGTTCAGATCAGAGCTCTGGGCGCGGAGTCCCA 2204
648 nPhePheHisLysAspSerThrLysThrArgArgGlyPheAlaGlnHisLeu 665
|||||.....:|||||.....:|||||.....:
2205 TTTCCTAGATTAAAGATAAAGAAAGGAAAGCAAAATCCCTCATTAAT 2254
665 eArgGlyTyrValIleGlyGlyAsnLeuHisThrCysSerAspLysIle 681
|||||.....:|||||.....:|||||.....:
2255 CTGGGATGATGCTATCGAGGTGCAAGCCCAACTGTCTTCAAAACTTA 2304
682 LeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPheVal 698
|||||.....:|||||.....:|||||.....:

```

```

2305 ATTAGCTTGGCTTTGGCCACTCTTTGGTAGCGATTAAGATTCTTACT 2354
698 1AlaLysAsnGlnGlyThrValTyrGlyGlyThrLeuTyrTyrGlnHisA 715
|||||.....:|||||.....:|||||.....:
2355 CGCTAAATAATCATATGATACCTATGACAGAGGCTCTTATATCCAAAC 2403
715 snGluThrTyrIleSerLeuProCysLysLeuArgProCysSerLeuSer 731
|||||.....:|||||.....:|||||.....:
2404 .....ATTACGAATGATAGTGGGTTCATAGTTGTCTTAGAT 2442
732 TyrValPro.....ThrGluIleProValLeuPheSerGlyAsnLe 745
|||||.....:|||||.....:|||||.....:
2443 AAACCTTCGTGCTCTGGAGTCATAAACCCCTGTTTGAAGGCGAGT 2492
745 uSerTyrThrHisThrAspAsnAspLeuLysThrLysTyrThrTyrP 762
|||||.....:|||||.....:|||||.....:
2493 CGGTTATACCCAGCTGATGAATGATCTGAAGCAAGTATCTGCGTATC 2542
762 rOThrValLysGlySerTrpGlyAsnAspSerPheAlaLeuGluPheGly 778
|||||.....:|||||.....:|||||.....:
2543 CTGAGGTGAAGAGTCTTGGGGAATATGCTTTTAACATGATGTTGGGA 2592
779 GlyArgAlaProIleCysLeuAspLeuSerAlaLeuPheGluGlnTyrMe 795
|||||.....:|||||.....:|||||.....:
2593 GCTTCTCTCATTTCTATATCCCTGAATACCTGATGTTTGAATACCTATGC 2642
795 tProPheMetLysLeuGlnPheValTyrAlaHisGlnGlnGlyPheLysG 812
|||||.....:|||||.....:|||||.....:
2643 TCCATATCATCAAACTGAATCTGACTATATATAGTCTGAGACAGCTTCTGG 2692
812 1uGlnGlyThrGluAlaArgGluPheGlySerSerArgLeuValAsnLeu 828
|||||.....:|||||.....:|||||.....:
2693 AGAAAGTACAGAGAGAGATCTTTGATGACAGACCTTCTCATATTTA 2742
829 AlaLeuProIleGlyIleArgPheAspLysGluSerAspCysGlnAspAl 845
|||||.....:|||||.....:|||||.....:
2743 TCTTATCCATAGGGGTGAGTGTGAGAAAGTCTCTGATTTATATGACTT 2792
845 aThrTyrAsnLeuThrLeuGlyTyrThrValAspLeuValArgSerAsn 862
|||||.....:|||||.....:|||||.....:
2793 TTCTATATGTCGACTTATATCCATATGCTTCGATCTTATCCGCATATGTC 2842
862 rOAspCysThrThrThrLeuArgIleSerGlyAspSerTrpLysThrPhe 878
|||||.....:|||||.....:|||||.....:
2843 CCAATGCACTACAGCACTGTATACAGCGAGCCCTTGGGAACCTTAT 2892
879 GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisP 895
|||||.....:|||||.....:|||||.....:
2893 GCCATATACCTTAGCAGCAGAGCCCTTGCAAGTGGTGCAGGCACTACTA 2942
895 eCysPheAsnSerAsnPheGluAlaPheSerGlnPheSerPheGluLeuA 912
|||||.....:|||||.....:|||||.....:
2943 CGCCTTCTCTCTATGTTGAAGTGTGCGCCACTTGTCTTGAATCTTC 2992
912 rGgGlySer 914
|||||.....:|||||.....:|||||.....:
2993 GTGGATCC 3000
seq_name: gb_baz:CPOMP54
seq_documentation block: 6030 bp DNA BCT 22-JAN-1999
LOCUS CPOMP54
DEFINITION Chlamydia pneumoniae omp5 and omp4 genes.
ACCESSION AJ001311
VERSION AJ001311.1 GI:3255934
KEYWORDS omp4 gene; omp5 gene; outer membrane protein 4; outer membrane protein 5
SOURCE Chlamydia pneumoniae.
ORGANISM Chlamydia pneumoniae
Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
REFERENCE 1 (bases 1 to 6030)
AUTHORS Knudsen,K.
TITLE Direct Submission

```

REFERENCE	TITLE	JOURNAL	MEDLINE	FEATURES
1	Submitted 129-AUG-1997 Knudsen K., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin building, University of Aarhus, DK-8000 Aarhus C, DENMARK 2 (bases 1 to 6030)	Knudsen, K., Madsen, A.S., Mygind, P., Christiansen, G. and Birkejund, S.	Identification of two novel genes encoding 97- to 99-kilodalton outer membrane proteins of <i>Chlamydia pneumoniae</i>	Infect. Immun. 67 (1), 375-383 (1999)
2	Location/Qualifiers			
	1..6030			
	/organism="Chlamydia pneumoniae"			
	/isolate="CDC/CWL-029/VR-1310"			
	/db_xref="taxon:83558"			
	/clone="BAP3"			
	/clone="pEX2-18"			
	/dev_stage="elementary body"			
	/clone_id="pLuescript"			
	/clone_id="pEX2"			
	complement(1..2697)			
	/gene="omp5"			
	/product="outer membrane protein 5"			
	complement<1..2742)			
	/gene="omp5"			
	/note="putative lipoprotein"			
	/codon_start=1			
	/transl_table=1			
	/product="outer membrane protein 5"			
	/protein_id="CAA04671.1"			
	/db_xref="GI:3255935"			
	/db_xref="SPTREMBL:O86163"			
	/translation="MKSQFWLVLSSTLACFTSCSTVFPAATENIGSDSFDGNTNGTYPKPNNTTGIDYLLTGDIPLQNLGSDAALTKGCFSTTBESLSPAGKGYSLSPFNIKS			
	SABEALSVTTDKNLSTLGTGSLFLPLAASSVITTPSGAKVKGCDLEPNNGTILPFG			
	KODCEGNGAISTKNLSLKNSTGSIPEEGKSSATGKGALCATGVDTNTATPPT			
	LFDSNNIEAAGAINSTNGCTITGNTSLVPSENVTAAGNGALSGDADVTISGNS			
	VTPSGNAVANGAITYAKKLLTAGCGGGISFSNNITVQGTAGNGAISTLAECSEL			
	SAEGDITFNGNAIVATTPOTTRKNSIDIGSTAKITLRLAISGHSIFFYDTITANTAA			
	DSDTLNLKADAGNSITDYSGSIVFSGEKLSEDAKAVADNLITSLKOPVLTAAQLAV			
	KRGVTLDTKGFTQAGSSVIMADAGTTAKASVEEYTLGLSLPYSLGEEKKVVYAASA			
	ASKNALVSGPLLDNGNAEYENDLKTODESFVOLSLGATITTVPAVPAVPTP			
	HYCGTGWMTGTVDDSTPTKTTATLANTGTGLPBERGGPIVPSNISSEFDIOAL			
	IOGVIERSLATLCSDRGFMAAGVANPILDKKKGKKRYRHRSGGYAIGAACOTSEML			
	ISFAFCOLFGSKDFLVAKNHTDTYACAFYIQHTTECSGFTGLDLPLPSGMRKPLY			
	LEGOLATSHVSNDLKTKYTAPEVKGSGMNAFMMLGASSHSYPEYLHCCDYLTAVY			
	KLNTVYIRDSFSEKGTGREGFSDSNLNFNLSPIGVFEKFSQDNDSDYDLTSLYSD			
	LIRNDPCTTALVYISGASWETYYANNLRAOLAVAGSHYAFSPMEFVILGQVFVEVRS			
	"			
gene	complement(1..2804)			
	/gene="omp5"			
	complement(12658..2742)			
	/gene="omp5"			
	complement(2747..2752)			
	/gene="omp5"			
	/note="putative"			
	complement(2777..2782)			
	/gene="omp5"			
	/note="putative"			
	complement(2799..2804)			
	/gene="omp5"			
	2863..5726			
	/gene="omp4"			
	2863..2868			
	/gene="omp4"			
	/note="putative"			
	2896..2902			
	/gene="omp4"			
	/note="putative"			
	2905..2955			
	/gene="omp4"			
	2905..5691			
	/gene="omp4"			
	"			
sig_peptide				
RBS				
-35_signal				
gene				
-10_signal				
-10_signal				
sig_peptide				
CDS				

```

/ncode="putative lipoprotein"
/transl_table=11
/product="outer membrane protein 4"
/protein_id="CA04672.1"
/db_xref="GI:325936"
/db_xref="SPTREMBL:086164"
/translation="MKTSIPMYVSVSVAFSCHTLOSLANEELL
TPKTSATITSLTGDFEFIEPKGKTPLSDCFOTDNLTLFLGK
GAASSTANKNLVTFSGFSLLSFDSSPTVTTVGGLTSAGGV
TADGALIGASGFLLTGTSGLAFENNSSSTKGALINTAGARR
TSGAIDDEGTSILSNKFLYFEFGNAARTGGALINTKAGSR
VAETSGGAIHAKKLALSSGGEFEPLRNWASATPKGALSIDA
VRNLTTGTSSTDPKRNAINIGSNCKFLNRAKHNTIFEPD
GSAALINPGOCSTLFECEFTLTADELKVDNKLSETPQVSLG
FSOAGSNLYMGDCSTLSTAGSITITTLGIVNDSLKQKPS
LNLIDISGNIYESHMFESHDQLESLKITVDADVTDVNDLSLI
QWNAVNTDIAATNKETAATATKTFVSPERSKALVCNTLWGL
AIGMHWKGTFWSSMTNLHKTTGDNRGFRHTSGGVYGGSA
FARPKDCEFIANNNSRTYGTGLFEKSHSLQOPNYRLGRKFS
DVQVFSHSDNRMEHTYSLPESEGSHTLCAAGSGLDLPV
KVEMYVONSFEFESSDGRGFSIGRLNLSIPVAKFYQGDJ
VRRNPQSTALVMSPDMSKIRGNLSQAFLRBSNNVYNSM
SRNVDTGKTRF"
2950..5688
/gene="omp4"
/product="outer membrane protein 4"
5698..5726
/gene="omp4"
/ncode="putative"
BASE COUNT 1760 a 1341 c 1195 g 1734 t
ORIGIN

alignment_scores:
    Quality: 2000.00
    Ratio: 2.801
    Percent Similarity: 76.282
    Percent Identity: 46.261

alignment_block:
US-09-428-122-2 x CPOMP54/rev ..

Align seg 1/1 to reverse of: CPOMP54 from: 1 to: 6030

1 MetLysSerSerPheProLysPheValphe...SerThrPheAlaIleph 16
|||||:||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:
2742 ATGAAATGCCAAATTTCTCTGTTAGTGCTCTCTCGACATTGGCATTTT 26933
16 e.....ProLysSerMetIleAlaThrGluThrValLeuSperS 30
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2692 TACTACTTTGTCACGCTTTTCTGCTGCAACTGCTGAATAATATAGCCCT 26433
30 eAlaSerPheAspGlyAsnLysAsn...GlyAsnPheSerValArgL 45
|||:|||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:
2642 CTGATAGCTTTTGCAGGAAGTACACACAGCACCTTACTCTCTAAAAAT 25933
46 SerGlnGluAspAlaGlyThrThrIleuPheLysGluValAsnIle 62
:::|||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:
2592 AGC.....ACTACTGGAATAGCATATACTCTGACAGAGATTAATCT 25494
62 uGluAsnIleProGlyThrGlyThrAlaIleThrIleLysSerCysPheAsn 79
|||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2548 GCAAAACCTT...GGGAGATCGGCAGACTTAAAGAAAGGATGTTTTCTG 25020
79 snThrIleGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
::|:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2501 ACACGTACGGAATCTTTAAAGCTTTGGCCGGTAGAGGATCACTCTTTCT 24520
96 GlnThrValAspAlaGlyThrValAlaGlyAlaValAlaAsnSerSer 112
:::||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:
2451 TTAAATATTAAGCT...AGTCGTGAAGCGCAGCACTT...TCTGTAC 24088
112 ValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIle 129

```

```

2407 AACTGATTAATAATCTGCGCTACAGGATTTTTCAGTCTTACTTTCTTAG 2358
129 laseProGlySerSerIleThr.....GlyGlyAlaIValSer 143
12357 CGGCCCATCATCGTATCAACACCCCTCAGGAAAGGTGCAGTTAA 2308
144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPhe 160
2307 TGTGGA...GGGATCTTACATTGATACCAATGAGACTATTATTATTA 2261
160 rLysAsnPheSerThrAspAsnGlyAlaIleThrAlaLysThrLeuS 177
2260 ACAAGATTACTGTGAGAAATGGCGAGCATTTCTACCAAGATCTT 2211
177 erLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSer 193
2210 CTTTGAAAAACACGCGGATTCCTTTTGAGGGAATAAATCGAGC 2161
194 .....LysLysGlyAlaIleGlnThrSerAspAlaLeuThrI 207
2160 GCACACGAGGAAAAAGTGGGCTATTGTCTACTGTACTGTAGATAT 2111
207 eThrGlyAsnGlnGlyGluValSerPheSerAspAsnThrSerSer 224
2110 TACAATTAATACGCTCCCTACCTCTCTGCAACAATATTGCTGAAG 2061
224 erGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAla 240
2060 CAGGTGAGCTATTAATAGCACAGAAACGTACTACAGGGAATACG 2011
241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerThr 257
2010 TCTCTGTATTCTTGAATAAGTGTACA.....GCGAGCC 1973
257 rGlyAspMetSerGlyGlyAlaIleCysAlaTyrLysThrSerTh 274
1972 AGGA.....AATGAGAGACTCT.....TCTGAGATG 1944
274 hrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsnThr 290
1943 CGCATGTTACCATATCGGATCGAATGAGTACTTCTCAGGAACAA 1894
291 SerThrThrAlaGlyAlaIleTyrValLysLysLeuLeuAlaSe 307
1893 GCTGTACCTAATGGCGAGCATTTATGCTAAGACCTTACACTGGCTT 1844
307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGly 322
1843 CGGGGGGGGGGGGTATCTCTTTCTTCACAATATAGTCCAAGGTACA 1794
322 hrLarProLysGlyAlaIleAlaIleGluAspSerGlyLuleSer 338
1793 CTCAGGTAATGtGGAGCATTTCTATCTGCGAGCTGAGAGTGTAGT 1744
339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValTh 355
1743 CTTTCACACAGACGAGGACATTACCTTCAATGGCAATGCCATTGTGC 1694
355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThSer 371
1693 AACTACACCAACAACATAAGAAATCTATGACATAGATCTACTG 1644
371 lalysMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleTyrPhe 387
1643 CAAGATCACAGAAATTACTGCAATATCTGGCATACATCTTTTCTAC 1594
388 AspProIleThrThrGlySerSerThrValThrAspValLeuLysVa 404
1593 GATCCGATTACTGTAATACGGCTGGGATTTCAACAGATCTTAAATCT 1544
404 lAsnGluThrProAlaAspSerAlaLeuGlnTyrThrGlyAsnIleLe 421
1543 CATAAAGCTGATGACGATTAATAGTACAGATTATAGTGGTGATTTGTT 1494
421 heThrGlyLulLysLeuSerGluThrGluAlaAlaAspSerLysAsn 437
1493 TTTCTGTGTAAGAGCTCTGGAAGATGAAGCAAAAGTTGCACACAAC 1444
438 ThrSerLysLeuLeuGlnProValThrLeuSerGlyLysThrLeuSe 454
1443 ACTTCACTACCTAGACAGCGCTGACTACTACTGACAGAAATTGTACT 1394
454 ulysHisGlyValThrLeuGlnThrGlnAlaPheThrGlnAlaAspS 471
1393 TAACTGTGTGCACCTCTGATCGAAGAGCTTACTCTGACCGCGGTT 1344
471 eArgLeuGlnMetAspValGlyThrThrLeuGluProAla...AspTh 486
1343 CTCTGTATTATGATGCGGCGACACCTTAAAGCAAGTACAGAGAG 1294
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAl 503
1293 GTCACTTTAACAGCTCTTCCATCTCTGATACCTTTAGCGGAGGTAA 1244
503 sLysAlaLysIleGluThrLysAlaIleThrSerLysAsnLeuThrLe 520
1243 GAAGTTGTAATTGCTGCTTCGCACAGCAAGTAAATGTAGCCCTTAG 1194
520 lYThrIleThrLeuLeuAspProThrGlyThrPheTyrGluAsnHis 536
1193 GTCCGATTTCTTTTGGATTACCAAGGAAATGCTTATGAAATACAGC 1144
537 LeuArgAsnProGlnSerTyrAspIleLeuGluLeuLysAlaSerGly 553
1143 TTAGCAAAACCAAGACTTTCATTTGTCAGCTCTGCTGGGTAC 1094
553 rValThrSerThrAlaValThrProAspProIleMetGlyLulLysP 570
1093 TGGACACACTAGAGATGTCACGCGTCTCTACAGTACACTCTACGC 1044
570 lstyGlyTyrGlnGlyThrTrpGlyProIleValTrp..... 582
1043 ACTATGGTATCAAGTACTTGGGA...ATGACTTGGTGTATGATACC 997
583 .....GlyThrGlyAlaSerThrThrAlaThrPheAsnTrpThrLys 598
996 GCAGAGACTCCAAAGACTAAGACAGGACATTAAGCTTGGACCAATAC 947
598 yTyrIleProAsnProGluArgIleGlySerLeuValProAsnSerLe 615
946 CTACCTTCGAATCCAGACGTCMAAGACCTTAGTCTCTAATAGCCTT 897
615 rPAsnAlaPheIleAspIleSerSerLeuHisTyrLeuMetGluThrAl 631
896 GGGGATCTTTTCAGCATCCCAAGCATTCMAAGTGTCTAAGAGAAAT 847
846 GCTTACACTTTTGTTCAGATCAGGCTCTGCGCGGAGTGCACCA 797
632 AsnGluGlyLeuGlnLysAspArgAlaPheTrpCysAlaGlyLeuSe 648
846 GCTTACACTTTTGTTCAGATCAGGCTCTGCGCGGAGTGCACCA 797
648 nPhePheHisLysAspSerThrLysThrArgArgGlyPheAlaGlnHis 665
796 TTTCTTAGTAAGAAATGAAGAGGGAAGCAAAATCCCTCATTAAT 747
665 erGlyLysTyrValIleGlyLysAsnLeuHisThrCysSerAspLysI 681
746 CTGGTGGAATGCTATCGAGGAGGCGCAAACTTGTTCTGAAACCTTA 697
682 LeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPhe 698
696 ATTAGCTTGGCTTTTCCACACTTGTGTAGCGATAAAGATTTCTTAAT 647
698 lAlaLysAsnGlnGlyThrValTyrGlyLysThrLeuTyrTyrGlnHis 715
646 CGCTAAAAATCATATGATACGATACAGAGACCTTCTATATCCAAAC 598

```



```

715 sngluThrTyrIleSerLeuProCysLysLeuArgProCysSerLeuSer 731
    |||::: |||::: |||:::
597 .....ATTACGAATGATGAGGTGCTATAGATTGCTCTTACAT 559
732 TyrValPro.....ThrGluLeuProValLeuPheSerGlyAsnLe 745
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
558 AAACCTCCGTGCTCTTGAGATCAATAACCCCTGTTTGAAGAGGCACCT 509
745 uSerTyrThrHisThrAspAsnAspLeuLysThrLysTyrThrThrTyr? 762
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
508 CGCTTAATGACGACGTCAGTATGATCTGAACACAAAGTATACTGCTATC 459
762 roThrValLysGlySerTrpGlyAsnAspSerPheAlaLeuGluPheGly 778
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
458 CTGAGGTGAAGGTTCTTGCGGGAATATGCTTTAATCATGATGTTGGGA 409
779 GlyArgAlaProIleCysLeuAspGluSerAlaLeuPheGluGlnTyrMe 795
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
408 GCTTCTCTCTTCTTATCCTGAATACCTGCATGTTGTTGATACCTATGC 359
795 tProPheMetLysLeuGlnPheValTyrAlaHisGlnGlnGlyPheLysG 812
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
358 TCATATACACTGAATCTGACCTATATAGTCAGACAGCAGCTTCTCG 309
812 IuGlnGlyThrGluAlaArgGluPheGlySerSerArgLeuValAsnLeu 828
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
308 AGAAGAGTACGAAGAGATCTTTGATGACACCAACCTCTCAATTTA 259
829 AlaLeuProIleGlyIleArgPheAspLysGluSerAspCysGlnAspAl 845
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
258 TCTTGGCTAATAGGGGTGAAGTTTGAGAGATTCCTGATTTAATGACTT 209
845 aThrTyrAsnLeuThrLeuGlyTyrThrValAspLeuValArgSerAsp 862
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
208 TTCTTATGATGTGACTTATCTTATGCTCTGATCTTATCCGCAATGATC 159
862 roAspCysThrThrLeuArgLysSerGlyAspSerTrpLysThrPhe 878
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
158 CCAATAGCTACGACCTGTATACAGCGGAGCCTCTTGGAACCTTAT 109
879 GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisPhe 895
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
108 GCCAATAGCTACGACGACGAGCCTTGCAAGTGGTGAGGCACTACTA 59
895 eCysPheAsnSerAsnPheGluAlaPheSerGlnPheSerPheGluLeu 912
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
58 CGCCTCTCTCTCTATGTTTGAAGTCTCGGCCCATTTGCTTTGAAGTTC 9
912 rGlyLysr 914
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
8 GTGATGCC 1
seq_name: gb_pat:A81839

```

```

seq_documentation_block: 2787 bp DNA PAT 21 JAN-2000
LOCUS A81839
DEFINITION Sequence 13 from Patent WO9858953.
ACCESSION A81839
VERSION A81839.1 GI:6731870
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2787)
AUTHORS Madsen,A. and Birkelund,S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL Patent: WO 9858953-A 30-DEC-1998;
MADSEN ANNA-SOFIE (DK); BIRKELUND SVEND (DK)
FEATURES
source
1.2787
/organism="unidentified"
/db_xref="taxon:32644"

```

```

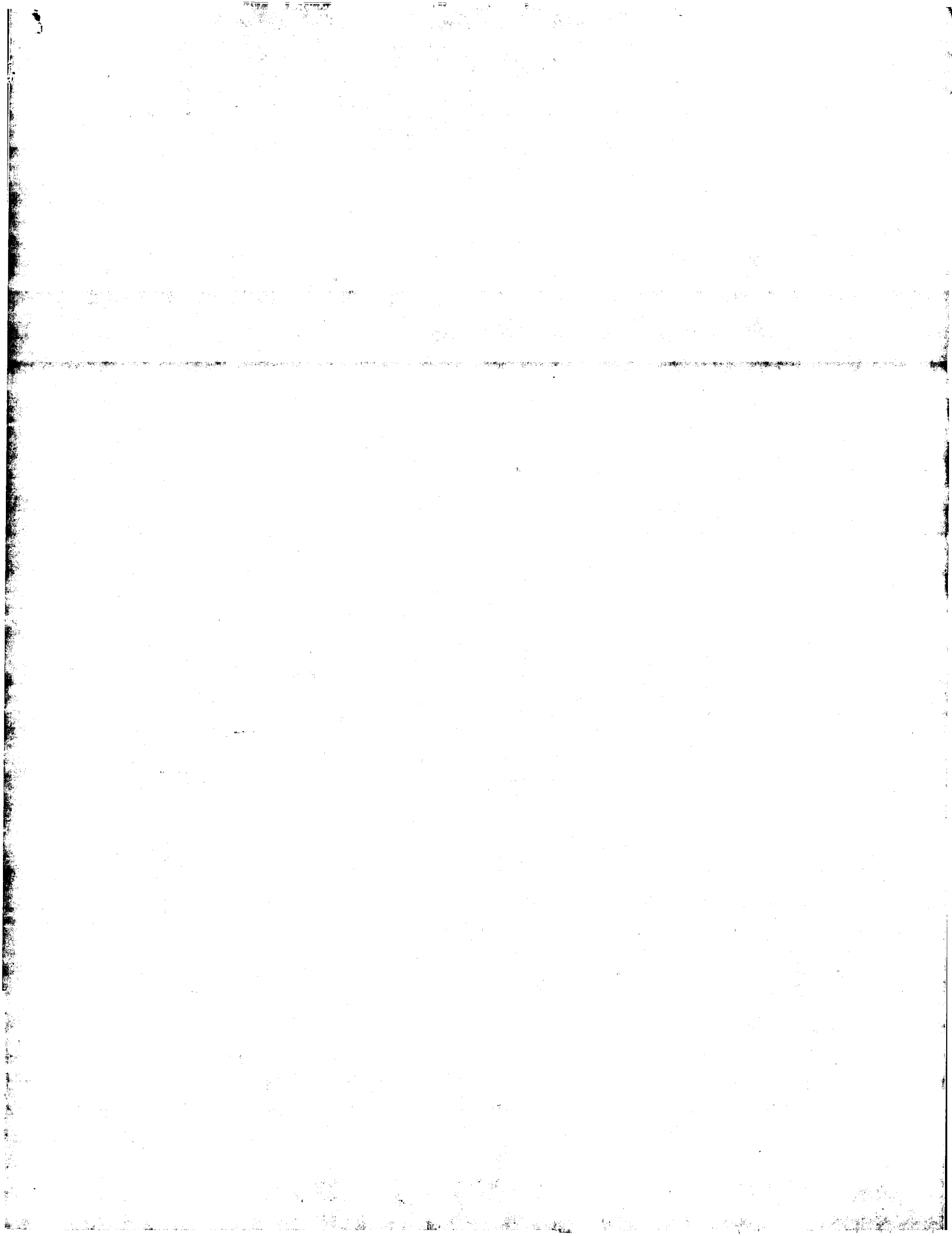
BASE COUNT 815 a 689 c 535 g 748 t
ORIGIN
alignment_scores
Quality: 1986.00 Length: 950
Ratio: 2.770 Gaps: 19
Percent Similarity: 75.474 Percent Identity: 45.263
alignment_block:
us-09-428-122-2 x A81839
Align seq 1/1 to: A81839 from: 1 to: 2787
1 MetLysSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1 ATGAATCTCTCTCTTCAATGTTGTTGTAATCTCGCATCTTAGACCTCC 50
17 oLeuSerMet.....IleAlaThrGluThrValLeu 28
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
51 CTGTCTCACTAATTTCTCTGCGTTTGCTGCTGTTGTTGAATCAATCTAG 100
28 spSerSerAlaSerPheAspGlyAsnLysAsnGlyAsnPheSerValArg 44
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
101 GACCTACCAATAGCTTCTCTGA.....CCAGAACCTACACTCCCA 144
45 GluSerGlnLysAla...GlyThrThrTyrLeuPheLysGlyAsnVal 60
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
145 GCCCAACAACAATGCGAGATGACATATCTAATATACAGAGGAGATGT 194
60 lThrLeuGlnAsnIleProGlyThrGlyThrAlaIleThrLysSerCysP 77
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
195 CTCAATACCAAT...GCAGATCTCCGACAGCTTAAACCGCTTCGCT 241
77 heAsnAsnThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeu 93
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
242 TTAAGAAGAACTAGCGGAATCTTCTTCCAGGCCAGCGCTACCAATT 291
94 LeuPheGlnThrValAspAlaGlyThrValAlaGlyAlaValAsnSe 110
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
292 CTCCTACAAATATCGATGCGGGA.....GCGAACGTACTCTTACCA 335
110 rSerValValAspLysSerThrThrPheIleGlyPheSerSerLeuSerP 127
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
336 TACAGCTCAATATAGCTTCTCTCTTTCAGAGATTCCTCATATTGTCAC 385
127 heIleAlaSerProGlySerSerIleThrThrGlyLysGlyAlaValSer 143
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
386 TAATACAAACC.....AGCAATGCTACACAGAGAACAGGACATCAAG 429
144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSe 160
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
430 ...TCCACAGAGACTGTTCTTATTCAGTCACTGCACTATGCTACTTGG 476
160 rLysAsnPheSerThrAspAsnGlyGlyAlaIleThrAlaLysThrLeuS 177
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
477 CCAAACTTTTCTATATGACATGAGAGCGCTCCCAAGGAGACTATCA 526
177 erLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer 193
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
527 GCTATGCG...CTAAACCCCAACCTTAACGTTGCCAAAACCAAGCAAG 573
194 LysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAs 210
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
574 CAAAAGGGGGTGCCTCTTTCACAGGAGGGAATTAACAATTAACATAC 623
210 nGlnGlyGluValSerPheSerAspAsnThrSerSerAspSerGlyAla 227
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
624 GTAAACTGACATCATTTCTGAAATATACGCGGGAACAATAGCGGAG 673
227 laIlePheThrGluAlaSerValThrIleSerAsnAsnAlaLysValSer 243
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
674 CCAATTAACAGGAAGCTAGCAGTTTATTATAGACGAACAACCAATTAGC 723

```



```

      :::::::::::::::::::: ||::: ||:::
2485 TCTGTGCTATTGGATTACCTTCGAGAGATTCTCGAGAAACGAGCGCTGC 2534
      845 aThrTYrAsnLeuThrLeuGlyTYrThrValAspLeuValArgSerAsnF 862
      :::::::::::::::::::: ||::: ||::: ||::: ||:::
2535 GTCTTAGAGAGCTACTGTCACTACGTTGCGGAGTGTATCGTAGAATG 2584
      862 rAspPCysThrThrThrLeuArgIleSerGlyAspSerTrpLysThrPhe 878
      :::::::::::::::::::: ||::: ||::: ||::: ||:::
2585 CTGACTGTCACGACGCTCTCTTAATCAACAATACCTCGTGAACAACTACA 2634
      879 GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisPh 895
      :::::::::::::::::::: ||::: ||::: ||::: ||:::
2635 GGAACGAATCTCTCAGACAAAGCTGTATCGAAGAGCAGGGATCTTTA 2684
      895 eCysPheAsnSerAsnPheGlnAlaPheSerGlnPheSerPheGlnLeuA 912
      :::::::::::::::::::: ||::: ||::: ||::: ||:::
2685 TGCCTTCTCTCAATCTTGAGGTACACAGTAACCTATCTATGSAATTC 2734
      912 rGcLYSerSerArgAsnTYrAsnValAspLeuGlyAlaLysTYrGlnPh 928
      :::::::::::::::::::: ||::: ||::: ||::: ||:::
2735 GTGGATCTTCACGACGCTACAAATGACAGATCTGGAGGTAAGTTCCAGTTC 2784
```



alignment_scores:

Quality: 4774.00 Length: 928
 Ratio: 5.144 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-428-122-2 x A27021 ..

Align seg 1/1 to: A27021 from: 1 to: 3000

```

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePheBr 17
  |||
101 ATGAAGTCTTCTTTCCCAAGTTGTATTTCTACATTGCTATTTTCCC 150
17 OLeuSerMetIleAlaThrGluThrValLeuAspSerSerAlaSerPheA 34
  |||
151 TTGTGCTATGATTGCTACCGAGACAGCTTTGGATTCAAGTGGAGATTGG 200
34 SPGLYAsnLysAsnGlyAsnPheSerValArgGluSerGlnGluAspAla 50
  |||
201 ATGGGAATATAAATGGTATTTTTCAGTTCGTGAGAGTCAGAGAGATGCT 250
51 GlyThrThrTyrLeuPheLysGlyAsnValThrLeuGluAsnIleProG 67
  |||
251 GGAAGTACTACTACTTTTAAGGAAATGTCACTCTAGAAAATATTCTGG 300
67 YThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspL 84
  |||
301 AACAGCGACAGCAATCACAAAAGCTGTTTAACACACTAAGGCGCAT 350
84 eutThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla 100
  |||
351 TGACTTTACAGGTACGGGAAGCTCTATTGTTCCAAACGCTGATGCA 400
101 GlyThrValAlaGlyAlaAlaValAsnSerSerValValAspLysSerTh 117
  |||
401 GGAAGTATGACAGGGGCGTCTGTTAACAGCAGCGGTAGATAATCTAC 450
117 rThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySers 134
  |||
451 CACGTTTATAGGGTTTCTCTGCTATCTTTATTGCGTCTCTGGAAGTT 500
134 erIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSer 150
  |||
501 CGATTAACACCGCGCAAGAGACCGCTTACGCTCTACGGGTAGCTTAGT 550
151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPheSerThrAspAs 167
  |||
551 TTGACAAAATAATGTCAGTTGCTCTCAGCAAAAACCTTTCAACGAGATA 600
167 nGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMetS 184
  |||
601 TGGGGGTGCTATCCCGCAAAAGCTCTTTCATTAAACGAGGACTACAATGT 650
184 erAlaLeuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGln 200
  |||
651 CAGCTCTGTTTCTGAAATAATACCTCTCAAGAAAGCGGAGCATTCAG 700
201 ThrSerSerPheAlaLeuThrIleThrGlyAsnGlnGlyGluValSerPhe 217
  |||
701 ACTTCCGATGCCCTTACCATTAAGTGAACCAAGGGGAAAGTCTCTTTTC 750
217 rAspAsnThrSerSerSerGlyAlaAlaIlePheThrGluAlaSerV 234
  |||
751 TGACAATACTTCTTCGATTCTGAGAGCTGCATTTTACAGAAAGCTCTCG 800
234 aIThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysValThr 250
  |||
801 TGACTATTCTATAATGCTAAAGTTCTTTATGACAAATAAGTCACACA 850
251 GlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaIleCysAl 267
  |||
851 GGAGCGAGCTCTCAACAAAGGGGATATGTCAAGAGGTGCTATCTGTGC 900

```

```

267 aTyrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
  |||
901 TTATAAAAGTGTACAGATACATAAGTCAACCTCTACCTGCAAAATCAGATGT 950
284 euleuPheSerAsnAsnThrSerThrAlaGlyAlaIleTyrVal 300
  |||
951 TACTCTTCAGCAACCAATACATCGACACAGCGGAGAGCTATCTATGTG 1000
301 LysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerThrArgnsE 317
  |||
1001 AAAAGCTCGACACTGCGCTCCGGAGAGACTTACCTTATTCAGTAGAAATG 1050
317 rValAsnGlyGlyThrAlaProLysGlyGlyAlaIleAlaIleGluAspS 334
  |||
1051 TGTCAAATGGAGGTACAGCTCTTAAGGTGAGGACCATAGCATGCAAGATA 1100
334 erGlyGluLeuSerLeuSerAlaAspSerGlyAspLysPheValPheLeuGly 350
  |||
1101 GTGGGGAATTGAGTTTATCCGCCGATAGTGGTGCATTTGCTTTTAAAGG 1150
351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
  |||
1151 AATAAGTCACTTCTACTACTCTCGGAGAGAAATAGAAAGTAGATGACTT 1200
367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAlaI 384
  |||
1201 AGGACAGAGTGCAAAGATGACACAGCTTGCGTTCTGCTGTGTAGAGCCA 1250
384 leTyrPheTyrAspProIleThrThrGlySerSerThrThrValThrAsp 400
  |||
1251 TCTACTTCATATGATCCCATACTACAGATCATCCAAACATTTTCAGAT 1300
401 ValLeuLysValAsnGluThrProAlaAspSerAlaLeuGlnTyrThrG 417
  |||
1301 GTCTTAAAGTTAATGAGACTCCGGCAGATTCTGCATCAATATATACAGG 1350
417 yAsnIleIlePheThrGlyGlyLysLeuSerLysThrGluAlaAlaAspS 434
  |||
1351 GAAACATCATCTTCACAGGAGAAAGTTATCAAGACAGAGGCGCGAGATT 1400
434 erLysAsnLeuThrSerLysLeuGlnProValThrLeuSerGlyGly 450
  |||
1401 CTAAAATCTTACTTGAAGTACTACAGCCTGTAACTCTTCAGAGAGT 1450
451 ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
  |||
1451 ACTTAATCTTAAACATGAGAGTGACTCTGCAGACTCAGGCATTCACATCA 1500
467 nGlnAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGluProA 484
  |||
1501 ACAGCAGATTCCTGCTCGAAATGAGAGTAGGAACACTCTACAGACTG 1550
484 lAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
  |||
1551 CTGATACTAGCACCATTAACATTTGGTCATTAACTACATGTTCTATAGAC 1600
501 GlyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuTh 517
  |||
1601 GGTGCAAGAGGCGCAAAATAGAAACCAAGCTACGTCAAAAATCTGCAC 1650
517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGluA 534
  |||
1651 TTTATCTGGAACCATCACTTATTTGACCCGACGCGGACGTTTATATAAA 1700
534 snHisSerLeuArgAsnProGlnSerTyrAspIleLeuGluLeuLysAla 550
  |||
1701 ATCATAGTTTAAAGAAATCCCTAGTCTACAGACATCTTAGAGCTCAAAAGCT 1750
551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGlyG 567
  |||
1751 TCTGGAAGTGTAAACACAGACCGCGAGTGAATCTCCAGATCTTAAATGGGTGA 1800

```

```

567 ulyspHeHisTyrGlyTyrGlnGlyThrTTPGLProIIeValTTPGLYT 584
|||||
1801 GAAATTCATTACGGCTATCAGGAGACCTGGGGCCAAATGTTGGGGGA 1850
584 hrGlyAlaSerThrThAlaThrPheAsnTrpThrLysThrGlyTyrIle 600
|||||
1851 CAGGGGCTTCTACGACCTGCACACTTCAACTGATAAAGCGGCTATATT 1900
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
|||||
1901 CCTAATCCCGAGCGTATCGGCTTTAGTCCCTAATAGCTTATGGAATGC 1950
617 aPheIIaSplIleSerSerLeuHisTyrLeuMetGluThrAlaSnGluG 634
|||||
1951 ATTATAGATATTAGCTCTCTCCATTCATTATGAGACGCAACAGAG 2000
634 lyleGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAspPhePhe 650
|||||
2001 GGTGCGAGGGAGACCGTCTTTGGTGCTGCTGATTATCAACTTCTTC 2050
651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyG 667
|||||
2051 CATAGAGATAGTACAAAACACGACGGGGTTTCGCCATTGAGTGGCGG 2100
667 YTYrValIIleGlyLysLeuHisThrCysSerAspLysIleLeuSerA 684
|||||
2101 TTATGCTATAGGAGAAACCTACATACTGTTCCAGATPACATCTTATG 2150
684 lAlaIaPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAla 700
|||||
2151 CTGCATTTTTCAGCTCTTGAGAGATAGAGACTACTTGTGAGCTAG 2200
701 AsnGlnGlyThrValIYrGlyGlyThrLeuTyrTyrGlnHisAsnGlu 717
|||||
2201 AATCAAGGTACAGCTACGAGGAACCTCTATTACAGCAACGAAG 2250
717 rTYrIIleSerLeuProCysLysLeuArgProCysSerLeuSerTyrVal 734
|||||
2251 CTATATCTCTCTCTGCAACACTACGCGCTTGTGCTGCTTATGTC 2300
734 roThrGluIIleProValLeuPheSerGlyAsnLeuSerTyrThrHis 750
|||||
2301 CTACAGAGATTCCTGTTCTCTTTCAGAAACCTTACTACACCAATAG 2350
751 AspAsnAspLeuLysThrLysTyrThrTyrProThrValLysGlySe 767
|||||
2351 GATACGATCTGAAACCAAGATATACACATATCTACTGTTAAGAGAG 2400
767 rTPGLYAsnAspSerPheAlaLeuGluPheGlyIYrArgAlaProIle 784
|||||
2401 CTGGGGGAATGATGTTCCGCTTAGAATTCGGTGAAGAGCTCCGATT 2450
784 yslLeuAspGluSerAlaLeuPheGluGlnTyrMetProPheMetLys 800
|||||
2451 GCTTAGATGAAAGGCTCTATTGAGCAGTACATGCCCTTCATGAAAT 2500
801 GluPheValIYrAlaHisGlnGluGlyPheLysGlnGlnGlyThrGlu 817
|||||
2501 CAGTTTCTATGACATCAGGAAGGTTTAAAGACAGGGAAGAGAGAGC 2550
817 aArgGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGly 834
|||||
2551 TCGTGAATTTGGAAGTACGCGCTGTGTGAATCTGCTTACCTTACGGA 2600
834 leaArgPheAspLysGluSerAspCysGlnAspAlaThrTyrAsnLeu 850
|||||
2601 TCCGATTGTGATGAAGATCAGACGCGCAAGATGCACAGTACAACTAC 2650
851 leuGlyTyrThrValAspLeuValArgSerAsnProAspCysThrTrp 867
|||||
2651 CTGCGTATTACTGCGATCTTGTTCGTAGTACCCGACGCTACGACAAAC 2700
867 rLeuArgIleSerGlyAspSerTyrLysThrPheGlyThrAsnLeuAla 884

```

```

|||||
2701 ACTGCGAATTAAGCGGTATCTTGAAACCTTCGGTACCAATTTGGCAA 2750
884 rGlnAlaLeuValIleuArgAlaGlyAsnHisPheCysPheAsnSerAsn 900
|||||
2751 GACAGCTTTAGTCTTCGTCGACGAGGACCATTTTTCCTTAACCTCAAT 2800
901 PheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSerArgas 917
|||||
2801 TTGGAACCTTTAGCCCAATTTCTTTGAATTCGTCGGTCATCTGCCAA 2850
917 nTYrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
|||||
2851 TTACAAATGTAGACTTAGGAGCAAAATACCAATTC 2884

seq_name: /SID6/gcdata/geneseq/geneseqn/NA1999.DAT:X06820
seq_documentation block:
ID X06820 standard; DNA: 2787 BP.
XX
AC X06820:
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp8 DNA.
XX
KW Omp8: outer membrane protein 8; surface exposed protein; antigen;
KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
OS Chlamydia pneumoniae.
XX
PN W09858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygind P;
XX
DR WPI: 1999-105610/09.
DR P-PSDB: W88421.
XX
PS Claim 6; Page 52-53; 115pp; English.
XX
CC This DNA sequence codes for the novel 90.0 kDa surface exposed
CC protein Omp8 (see W88421) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see X06816-27) encoding
CC Omp4-Omp15 proteins (see W88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89,6-100,3 kDa and
CC about 56,1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA

```

CC vaccines for effecting in vivo expression of antiens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.

50 Sequence 2787 BP; 811 A; 583 C; 598 G; 795 T; 0 other;

alignment_scores:

Quality: 4760.00 Length: 928
Ratio: 5.135 Gaps: 0
Percent Similarity: 99.892 Percent Identity: 99.677

alignment_block:

US-09-428-122-2 x X06820 ..

Align seg 1/1 to: X06820 from: 1 to: 2787

```
1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePheR 17
|||||
1 ATGAAAGCTCTCTCCCAAGTTGTATTCTTACATTGCTATTTCC 50
17 oleuSerMetIleAlaThrGluThrValleuAspSerSerAlaSerPhe 34
|||||
51 TTGTGTATGATTGCTACCGACAGATTGGATTCAAGTGGAGTTTCG 100
34 spcLysnLysAsnGlyAsnPheSerValArgLysSerGlnLysAla 50
|||||
101 ATGGGAATAAAATGGTAATTTTTCAGTTGCTGAGATCGAGAAATGCT 150
51 GlyThrThrTyrIleuPheLysGlyAsnValThrLeuGlnAsnIleProG 67
|||||
151 GGAACTACCTACCTATTTAAGGAAATGTCACCTAGAAAATTTCTCTG 200
67 yThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAsp 84
|||||
201 AACAGGCACAGCAATCACAAAAGCTGTTTAAACACACTAAGGGCGATT 250
84 euThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla 100
|||||
251 TGACTTTCACAGGTACGGGAAGCTCTATTTGTTCCAAAGGTGATGCA 300
101 GlyThrValAlaGlyAlaValAsnSerSerValValAspLysSerThr 117
|||||
301 GGGACTGTAGCAGGGGCTGCTGTTACACAGCGGTGTAGATAAATCTAC 350
117 rThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySer 134
|||||
351 CACGTTTATAGGGTTTCTTCGCTATCTTTATTCGCTCTCCGGAAGTT 400
134 erLleThrThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSer 150
|||||
401 CCATAACTACCGCAAGAGAGCGGTAGCTGCTACGGGTAGCTTGAAAG 450
151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPheSerThrAsp 167
|||||
451 TTGTGACAAAATGTCAAGTTGCTCTTCACAAAACCTTTTCAACGGATA 500
167 nGlyGlyAlaIleThrAlaLysThrPheSerLeuThrGlyThrMet 184
|||||
501 TGGCGGTGTATCACCAGCAAAACCTTTTCATTACAGGAGCTACAAAT 550
184 erAlaLeuPheSerGlnAsnThrSerSerLysGlyGlyAlaIleGln 200
|||||
551 CAGCTCTGTTTCTGAAAATACCTCTCAAGAAAGCGAGCATTCAG 600
201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyLysValSerPhe 217
|||||
601 ACTTCGAGAGCCCTTACCACTTACGAAACCAAGGGAGAGTCTCTTTTC 650
217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGlnLysVal 234
|||||
651 TGACAAATCTCTTCGGAATTCGAGCTGCATATTTTACAGAAAGCTCGG 700
```

```
234 alThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysValThr 250
|||||
701 TGACTATTCTATAATATGCTAAAGTTTCCCTTATTGACATATAGGTACA 750
251 GlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaIleCysAl 267
|||||
751 GGAGCGAGCTCCTCAACAACGGGGGATATGTCAGAGGTCTATCTGTGC 800
267 aTyrlYsThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMet 284
|||||
801 TTATAAACTAGTACGATACATTAAGGTACCTCCTCAGTGAATTCAGATGT 850
284 euLeuPheSerAsnAsnThrSerThrThrAlaGlyAlaIleTyrVal 300
|||||
851 TACTCTTCAGCAACAATATACATGACACAGCGGAGGAGCTATCTATGTC 900
301 LysLysLeuGlnLysAlaSerGlyGlyLeuThrLeuPheSerArgAsn 317
|||||
901 AAAAAGCTCGAACTGGCTCCGAGAGACTTACCTATTCAGTGAATAAG 950
317 rValAsnGlyGlyThrAlaProLysGlyGlyAlaIleAlaIleGlnAsp 334
|||||
951 TGTCAATGAGGTACAGCTCTTAAAGGTGAGCCATAGCTATCGAAGATA 1000
334 erGlyGlnLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
|||||
1001 GTGGGGAATGAGTTTATCCCGCATAGTGATGACATTGCTTTTAAAG 1050
351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAsp 367
|||||
1051 AATACAGTACTCTTACTACTCCTCGGAGCAATAGATAGTATGACTT 1100
367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleAlaIleGly 384
|||||
1101 AGGAACGAGTCAAGAGATGACAGCTTGCGTTCTGCTGCTGTTAAGCCA 1150
384 leryrPheTyrAspProIleThrThrGlySerSerThrThrValThrAsp 400
|||||
1151 TCTACTTCTATGATCCCACTAATACAGATCTTCCACAACAGTTACAGAT 1200
401 ValLeuLysValAsnGlnThrProLysAspSerAlaLeuGlnTyrThrG 417
|||||
1201 GTCTTAAAGTTAATAGAGCTCCGGCAGATTCTGCATCAATATACAGG 1250
417 yAsnIleIlePheThrGlyGlyLysLeuSerGlnThrGlnAlaIleAsp 434
|||||
1251 GAACATCATCTTCAAGAGGAAAGTTATCAGAGCACAAGCCGCGAGATT 1300
434 erLysAsnLeuThrSerLysLeuGlnProValThrLeuSerGlyGly 450
|||||
1301 CTAAAAATCTTACTTCGAAGCTACTACAGCCGTGTAACCTTTACAGAGGT 1350
451 ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaIlePheThrG 467
|||||
1351 ACTCTATCTTTAAACATGAGGTGACTCTGCAGACTCAGGATTCACATCA 1400
467 nGlnAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuGlnPro 484
|||||
1401 ACAGCAGATTCTGCTCTCGAAATGAGACTAGGAACCTCTATGAACTG 1450
1401 aAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
|||||
1451 CTGATACTAGCACCAATAAACATTTGGTCATTAACATCAGTTCTATAGAC 1500
501 GlyAlaLysLysAlaLysIleGlnThrLysAlaThrSerLysAsnLeuThr 517
|||||
1501 GGTGCAAGAGGCAAAATAGAAACCAAGCTACGTAATAAAATTCGAC 1550
517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGln 534
|||||
1551 TTTACTCTGAACCATCACTTATTGAGCCGACGCGACGTTTATGAA 1600
534 snHisSerLeuArgAsnProGlnSerTyrAspIleLeuGlnLeuLysAla 550
```



```

|||||
1601 ATCATAGTTTAAAGAAATCCTCAGTCTACGACATCTTAGAGCTCAACCT 1650
551 SerGlyThrValThrSerThrAlaValThrProAspProIleMeGlyG1 567
1651 TCTGGAACTGTACAGACACCCGACGTGACTGCAGATCCTATATATGGGGA 1700
567 uLysPheHisTyrGlyTyrGlnGlyThrTrpGlyProIleValTrpGlyT 584
1701 GAAATTCATTACGGCTATCAGGGAACCTGGGGCCCAATTGTTGGGGA 1750
584 hngLYALSerThrThrAlaThrPheAsnTrpThrLysThrGlyTyrIle 600
1751 CAGGGGCTTTACGACTCAACCTCAACTGGACTAAACCTGGCTATATT 1800
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
1801 CCTATCCCGAGCGTATCGGCTCTTAGTCCCTATAGCTTATGGAATGC 1850
617 aPheIleAspIleSerSerLeuHisTyrLeuMetGluThrAlaAsnGluG 634
1851 ATTATAGATATTACCTCTCCATTATCTATGAGAGCTGCAACGAGAG 1900
634 lYLeuG1G1YAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhePhe 650
1901 GGTTCCAGGAGACCGCTCTTTTGTGTGCTGATTAATCTACTTCTTC 1950
651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyG1 667
1951 CATAGAGTAGTACAAACACGACGCGGTTCCCATTTGAGTGGCGG 2000
667 YTrValIleGlyGlyAsnLeuHisThrCysSerAspLysIleLeuSerA 684
2001 TTATGTCATAGAGGAAACCTACATCTGTTCAGATTAAGATTCTTAGG 2050
684 lAlaLAPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAlaLys 700
2051 CTGCATTTTGTACGCTCTTGGAAAGATAGAGACTACTTGTACTAG 2100
701 AsnGlnGlyThrValTyrGlyGlyThrLeuTyrTyrGlnHisAsnGluTh 717
2101 AATCAAGGTACAGTCTACGAGGAACTCTCTATACCAACGAAACGAAAC 2150
717 rTyrlleSerLeuProCysLysLeuArgProCysSerLeuSerTyrValP 734
2151 CTATATCTCTCTCTCTGCAACCTAGCGCTGTGCTGTATGTC 2200
734 roThrGluIleProValLeuPheSerGlyAsnLeuSerTyrThrHisThr 750
2201 CTACAGAGATTCTCTCTCTTTCAGGAAACCTTAGCTACACCATACG 2250
751 AspAsnAspLeuLysThrLysTyrThrTyrProThrValLysGlySe 767
2251 GATAACGATCTGAAACCAAGATATACACATATCTACTGTTAAAGAGAG 2300
767 rTrpGlyAsnAspSerPheAlaLeuGluPheGlyGlyArgAlaProIleC 784
2301 CTGGGGGATGATAGTTCCGCTTAGAATTCCGTTGAAGAGCTCCGATTT 2350
784 ySLeAspGluSerAlaLeuPheGluGlnTyrMetProPheMetLysLeu 800
2351 GCTTAGATGAAGAGCTCTATTTAGAGAGTACATCCCTCATGAAATGG 2400
801 GlnPheValTyrAlaHisGlnGlnGlyPheLysGlnGlnGlyThrGluAl 817
2401 CAGTTGCTATAGCACATCAGAAAGTTTAAAGAACGAGGAACGAAAGC 2450
817 aArgGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGlyI 834
2451 TCGTAAATTGGAAGTACCGCTTGTGATCTTCCCTTACTATACGGGA 2500
834 lArgPheAspLysGluSerAspCysGlnAspAlaThrTyrAsnLeuThr 850
|||||

```

```

2501 TCCGATTGTGATAGGAATCAGACTGCCAAGATGCCAAGTACATCTACT 2550
851 LeuGlyTyrThrValAspLeuValArgSerAsnProAspCysThrThrTh 867
2551 CTGTGTTACTGTGATCTTGTCTGTAGTACCCGACGTCTAGACAC 2600
867 rLeuArgIleSerGlyAspSerTrpLysThrPheGlyThrAsnLeuAla 884
2601 ACTGGAATTAGCGGTGATTTCTTGAAACCTTCGTACGAATTTGGCA 2650
884 rGlnAlaLeuValLeuArgAlaGlyAsnHisPheCysPheAsnSerAsn 900
2651 GACAGCTTTTAGTCTTCGTGACAGGAAACATTTTGTCTTAACCTCAAT 2700
901 PheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSerArgAs 917
2701 TTGAGAGCTTTAGCCAAATTTCTTTGAATGGGTGGTATCTCGCA 2750
917 nTyraSnValAspLeuGlyAlaLysTyrGlnPhe 928
2751 TTACAATGTAGACTTAGAGCAAAATACCAATTC 2784
seq_name: /SID6/gcgdata/geneseq/geneseqn/NA1999.DAT.X91990
seq_documentation block:
ID X91990 standard: DNA; 1230025 BP.
AC
XX
AC X91990;
DT
XX
DT 13-SEP-1999 (first entry)
DE
XX Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
XX
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
XX vaccine; neutralising epitope; ss.
XX
XX Chlamydia pneumoniae.
XX
XX WO927105-A2.
XX
XX PD 03-JUN-1999.
XX
XX PF 20-NOV-1998: 98WO-IB01890.
XX
XX PR 04-NOV-1998: 98US-0107078.
XX
XX PR 21-NOV-1997: 97FR-0014673.
XX
XX PA (GEST ) GENSET.
XX
XX PI Griffiths R;
XX
XX DR WPI: 1999-357842/30.
XX
XX PT Genome sequence of Chlamydia pneumoniae
XX
XX PS Claim 1; Page 291-611; 1912pp; English.
XX
XX CC The present sequence represents the complete genome of Chlamydia
XX pneumoniae, and encodes proteins Y34584-Y35879. C. pneumoniae causes
XX respiratory disease such as pneumonia and bronchitis and is thought
XX to be a contributing factor in heart disease, sarcoidosis, sinusitis,
XX purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
XX encoded by the open reading frames of the C. pneumoniae genome (see
XX Y34584-Y35879) can be used in immunogenic compositions as vaccines.
XX Vectors containing C. pneumoniae nucleotide sequences can also be
XX used as immunogenic compositions, especially where the vector directs
XX the expression of a neutralising epitope of C. pneumoniae.
XX
XX SO Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;
alignment_scores:

```

Quality: 4733.00 Length: 929
 Ratio: 5.100 Gaps: 1
 Percent Similarity: 99.892 Percent Identity: 99.892
 Alignment_block:
 us-09-428-122-2 x X91990 ..
 Align seg 1/1 to: X91990 from: 1 to: 1230025

1 MetLysSerPheProLysPheValPheSerThrPheAlaIlePheP 17
 |||||
 31758 ATCAAGTCTTCTTCCCAAGTTGTATTTTCATATTGCTATTTC 31807
 17 OLeuSerMetIleAlaThrGluThrValLeuAspSerSerAlaSerPheA 34
 |||||
 31808 TTGTGTATGATGTCTACGAGACAGTTTGGATTTCAGATGGCAGTTTG 31857
 34 SPGLysAsnLysAsnGlyAsnPhSerValArgGluSerGlnGluAspAla 50
 |||||
 31858 ATGGGAATGAAATGGTAATTTTTCAGTTCGTGAGAGTCAGAGATGCT 31907
 51 GLyThrThrTyrLeuPheLysGlyAsnValThrLeuGluAsnIleProG 67
 |||||
 31908 GGAACTACCTACTATTTAAGGAAATGTCACTCTAGAAAATATTCTGG 31957
 67 YThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspL 84
 |||||
 31958 AACAGCGACAGCAATCACAAAAGCTGTTTAAACAACACTAGGGCGANT 32007
 84 eUThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAla 100
 |||||
 32008 TGACTTTCACAGGTAAAGGAACTCTCTATTGTTCCTCAACGGTGGATGCA 32057
 101 GLyThrValAlaGlyAlaAlaValAsnSerSerValAlaSpLysSerTh 117
 |||||
 32058 GGGACTGTAGCAGGGGCTGCTGTAAACAGACGCGTGAAGATTAATCTAC 32107
 117 ThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlySers 134
 |||||
 32108 CACGTTTATAGGTTTCTTCGCTATCTTTATTTGCGTCTCGTGAAGTT 32157
 134 eTleThrThrGlyLysGlyAlaValAsnSerSerThrGlySerLeuSer 150
 |||||
 32158 CGATTACTACCGGCAAGAGAGCCGTACTGCTACAGGGTAGCTTGAGT 32207
 151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPhSerThrAspAs 167
 |||||
 32208 TTGAC. AAAAATGTCAGTTGCTCTTCAGCAAAAACCTTTCACGGAGTAA 32256
 167 nGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMetS 184
 |||||
 32257 TGGGGTGCTATCAACGCCAAAACCTTCTTACACAGGAGCTACATGT 32306
 184 eAlaLeuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGln 200
 |||||
 32307 CAGCTCGTTTCTGAAAATACCTCTCAAGAAAGGGGAGCCATTGAG 32356
 201 ThrSerSpAlaLeuThrIleThrGlyAsnGlnGlyGluValSerPheSe 217
 |||||
 32357 ACTTCGATGCCCTTACCATTTACTGGAACCAAGGAGGAGCTCTTTTTC 32406
 217 rAspAsnThrSerSerAspSerGlyAlaAlaIlePheThrGluLasev 234
 |||||
 32407 TGACAATACTCTTCGGATTCTGGAGCTGCATTTTACAGAACCTTCGG 32456
 234 aLThrIleSerAsnAlaLysValSerPheIleAspAsnLysValThr 250
 |||||
 32457 TGACTATTTCATAATAGCTAAAGTTTCTTTATTGACAAATAAGTCACA 32506
 251 GLyAlaSerSerThrThrGlyAspMetSerGlyGlyAlaIleCysAl 267
 |||||
 32507 GGAGCGAGCTCTCAACAACGGGGGATATGTCAAGAGTCTATCTGTGC 32556

267 aTyrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
 |||||
 32557 TTATAAACTAGTACAGATACTAAGGTACACCCCTCAGCTGAATATAGATGT 32606
 284 eUleuPheSerAsnAsnThrSerThrThrAlaGlyGlyAlaIleTyrVal 300
 |||||
 32607 TACTCTCAGCAACCAATACATGACACAGCGGAGGAGGATCTATCTATGTG 32656
 301 LysLysLeuGluLeuAlaSerGlyGlyLeuThrLeuPheSerArgAsnSe 317
 |||||
 32657 AAAAAGCTCGAAGCGGCTCCGGAGGACTTACCTATTCCAGTGAATAATAG 32706
 317 rValAsnGlyGlyThrAlaProLysGlyGlyAlaIleAlaIleGluAspS 334
 |||||
 32707 TGTCAATGGAGGTACACCTCTTAAGGTGAGCCATAGCTATCGAAGATA 32756
 334 eRGlyGluLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGly 350
 |||||
 32757 GTGGGGAATTAGTTTATCCGCCGATAGTGTCATGTCTTTTAAAGG 32806
 351 AsnThrValThrSerThrThrProGlyThrAsnArgSerSerIleAspLe 367
 |||||
 32807 AATACAGTCACTTCTACTACTCTCGGGACGAATAGAAGTATGATCGACTT 32856
 367 uGlyThrSerAlaLysMetThrAlaLeuArgSerAlaIleAlaIleArgAla 384
 |||||
 32857 AGGACAGAGTGCAAAGATGACAGCTTGTGCTGTGCTGATGAGGCA 32906
 384 lEryPheTyrAspProIleThrThrGlySerSerThrThrValThrAsp 400
 |||||
 32907 TCTACTCTATGATCCCATTAAGTACAGAGATCATCCAAACAGTTCACAGT 32956
 401 ValLeuLysValAsnGluThrProAlaAspSerAlaLeuGlnTyrThrG 417
 |||||
 32957 GTCTTAAGTTATATGAGACTCGGCGAGATTCTGCTCAATATACAGG 33006
 417 YAsnIleIlePheThrGlyGluLysLeuSerGluThrGluAlaAlaAspS 434
 |||||
 33007 GAACATCATCTTCACAGAGAAAGTTATCAGAGACAGAGCGCCGAGATT 33056
 434 eTyrAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGlyGly 450
 |||||
 33057 CTAAAAATCTTACTTGAAGCTACTACACCCGTGAACCTTTTCAGAGAGT 33106
 451 ThrLeuSerLeuLysHisGlyValThrLeuGlnThrGlnAlaPheThrG 467
 |||||
 33107 ACTCTATCTTTAAACATGAGATGACTGTCAGACTCAGAGCTTACTCTCA 33156
 467 nGlnAlaAspSerArgLeuGluMetAspValGlyThrThrLeuGluProA 484
 |||||
 33157 ACAGCGAGATTCTCGTCTCGAAATGAGAGTACGAACTACTAGAACTG 33206
 484 lAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSerIleAsp 500
 |||||
 33207 CTGATCTGACCAACATTAACAAATTTGGTCATTAAACATAGTTCTATAGC 33256
 501 GLyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuTh 517
 |||||
 33257 GGTGCAAAGAGCAAAAATAGAAACCAAGCTAGCTCAAAAAATCTGAC 33306
 517 rLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGluA 534
 |||||
 33307 TTTATCTGGAAACCATCACTTTATTGGACCCACGCGGACGTTTATGAA 33356
 534 snHisSerLeuArgAsnProGlnSerTyrAspIleLeuGluLeuLysAla 550
 |||||
 33357 ATCATAGTTTAAGAAATCTCAGTCTACGCATCTTAAGACTCAAAAGCT 33406
 551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGlyG 567
 |||||
 33407 TCTGGAAGCTGTAAACAGCACCGCAGTGTACTCCAGATCTTAATAGGGTGA 33456
 567 uLysPheHisTyrGlyTyrGlnGlyThrThrProGlyProIleValThrPro 584

```

33457 GAAATTCATTACGGCTATCAGGAACTGGGCCCATTTGTTGGGGA 33506
584 hrc1yalaSerThrThrAlaThrPheasnTrpHlyThrGlyTyrIle 600
33507 CAGGGCTTACGACGCAACTTCACTGACATAAAGTGGCTATATT 33556
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
33557 CCTAAATCCACGAGGTATCGGCTTTAGTCCCTAATAGCTTATGGAATG 33606
617 aPheIleAspIleSerSerLeuHlyThrLeuMetGluThrAlaAsnGluG 634
33607 ATTATATGATATTAAGTCTCTCCATTAATCTTATGAGACTGCAACGAG 33656
634 LysGluGlnGlyAspArgAlaPheTrpCysAlaGlyLeuSerAsnPhphe 650
33657 GGTTCGAGGAGACCGTGGCTTTTGGTGGCTGATTAATCTTCTTC 33706
651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyG 667
33707 CATTAAGATATGATCAAAACACACGCGGGTTCCGCAATTTGAGTGGCG 33756
667 YTYrValIleGlyGlyAsnLeuHlyThrCysSerAspLysIleLeuSerA 684
33757 TTTATGTCATAGGAGAACTACATACTTGTTCAGATAAATCTTATG 33806
684 IAlaIlePheCysGlnLeuPheGlyArgAspArgAspTyrPheValAlaLys 700
33807 CTCATATTTGTGACGCTCTTGGAAAGATAGAGACTACTTTAGTAAAG 33856
701 AsnGln GlyThrValTyrGlyTyrLeuTyrTyrGlnHisAsnGluTr 717
33857 AATCAAGGATAGTACTACGAGAACTCTCTATTACACACACGAA 33906
717 hTrpIleSerLeuProCysLysLeuArgProCysSerLeuSerTyrVal 733
33907 CCTATATCTCTCTCTCTGCAAACTACGGCTTGTCTTATGTT 33956
734 ProThrGluIleProValLeuPheSerGlyAsnLeuSerTyrThrHisTh 750
33957 CCTACAGAGATTCCTGTTCTTTTCAGAAACCTTAGTACACCCATAC 34006
750 fAspAsnAspLeuLysThrLysTyrThrTyrProThrValLysGlyS 767
34007 GGATTAACGATCTGAAAACCAAGTATACACATATCTTACTGTTAAAGAA 34056
767 eTrpGlyAsnAspSerPheAlaLeuGluPheGlyArgAlaProIle 783
34057 GCTGGGGGATGATAGTTTCGCTTAGAATTCGGTGAACAGCTCCGATT 34106
784 CysLeuAspGluSerAlaLeuPheGluGlnTyrMetProPheMetLysLe 800
34107 TCGTTGATGTAAGAGTCTCTATTGAGCACTACATGCCCTTCATGAAT 34156
800 uGlnPheValTyrAlaHisGlnGluGlyPheLysGlnGlyThrGluA 817
34157 GCAGTTTGTATGACATACGAGAGTTTAAAGAACAGGACAGAG 34206
817 IaArgGluPheGlySerSerArgLeuValAsnLeuAlaLeuProIleGly 833
34207 CTCGTGAATTTGGAAGTAGCGCTTGTGAATCTTGCCTTACTATGGG 34256
834 IleArgPheAspLysGluSerAspCysGlnAspAlaThrTyrAsnLeuTh 850
34257 ATCCGATTGTGATAGGAATCAGACTGCAGAGATGCAACGTACATCTTAC 34306
850 rLeuGlyTyrThrValAspLeuValArgSerAsnProAspCysThrThyT 867
34307 TCTTGGTATACGTGTGATCTTGTCTGATTAACCCGAGATGTACAGAC 34356
867 hLeuValrGlyLeuSerGlyAspSerTyrPheGlyThrPheGlyThrAsnLeuAla 883

```

```

34357 CACTGCCAATTACGGTGATTTCTTGAAAACTTCCGTTACGAATTTGGCA 34406
884 ArgGlnAlaLeuValLeuValArgAlaGlyAsnHisPheCysPheAsnSerAs 900
34407 AGACAGCTTTAGTCTCTGTCGAGGAAACATTTTGTCTTATACACAA 34456
900 nPheGluAlaPheSerGlnPheSerPheGluLeuArgGlySerSerArgA 917
34457 TTTTGAAGCCTTACCAATTTTCTTGAATTCGCTGGGTCACTCGCA 34506
917 sPTrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
34507 ATTACATGTAGACTTATGAGAGCAAAATGCCAATTC 34541

seq_name: /SID56/gcdata/geneseq/geneseq/NA1999.DAT:X06817
seq_documentation_block:
ID X06817 standard; DNA: 2815 BP.
XX
AC X06817:
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp5 DNA.
XX
KM Omp5; outer membrane protein 5; surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX Chlamydia pneumoniae.
OS
FH Key 1.2787 Location/Qualifiers
FT CDS /*tag= a
XX
PN W09858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRKELUND S.
XX (CHRIL/) CHRISTIANSEN G.
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Mygind P;
XX WPI: 1999-105610/09.
XX P-PSDB: W88418.
XX
PT Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins
XX
PS Claim 6; Page 42-43; 115pp; English.
XX
CC This DNA sequence codes for the novel 97.2 kDa surface exposed
CC protein Omp5 (see W88418) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see X06816-27) encoding
CC Omp4-Omp13 proteins (see W88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used

```

CC in the diagnosis of *C. pneumoniae* infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting *in vivo* expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,
 CC which are possibly associated with *C. pneumoniae*.

XX Sequence 2815 BP; 790 A; 575 C; 626 G; 824 T; 0 other;

alignment_scores:

Quality: 2058.00 Length: 950
 Ratio: 2.831 Gaps: 20
 Percent Similarity: 76.526 Percent Identity: 46.737

alignment_block:

US-09-428-122-2 x X06817

Align seg 1/1 to: X06817 from: 1 to: 2815

```

1 MetLysSerSerPheProLysPheValPhe...SerThrPheAlaIlePh 16
  |||||
1 ATGAATTCGCAATTTCTCGTAGTCTCTCTCGACATTGGCATGTTT 50
  |||||
16 e.....ProLysSerMetIleAthrGluThrValLeuAspSers 30
  |||||
51 TACTAGTGTTCACACTGTTTTCCTGCACTGCTGAATAATATAGCCCTT 100
  |||||
30 eAlaSerPheAspGlyAsnLysAsn...GlyAsnPheSerValArgGlu 45
  |||||
101 CTGATAGCTTTGACGGAGTACTACACAGCAGCATTACTCTTAATAAT 150
  |||||
46 SerGluGluAspAlaGlyThrThrTyrrLeuPheLysGlyAsnValThrIe 62
  |||||
151 ACG.....ACTACTGGAATAGACTATACTGTGACAGAGATATACTCT 194
  |||||
62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsnA 79
  |||||
195 GCAAAACCTT...GGGGAATTCGGACACTTAAACGAGGGTGTGTTTTCG 241
  |||||
79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
  |||||
242 ACACATCGGAATCTTTAGCTTGGCGGTAAAGGGTACTCATTCTTTT 291
  |||||
96 GluThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSerSera 112
  |||||
292 TTAATATTAAAGTCT...AGTGCTGAAGCGCAGCAGCTT...TCTGTAC 335
  |||||
112 lValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleA 129
  |||||
336 AACTGATTAATAATCTGTCCTACACAGATTTTCGAGCTTACTTTCTTGG 385
  |||||
129 lAserProGlySerSerIleThrThr.....GlyLysGlyAlaValSer 143
  |||||
386 CGGCCCATCATCGGTAATCACACACCCCTCAGAAAAGGTGACAGTTAA 435
  |||||
144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSe 160
  |||||
436 TGTGGA...GGGATCTTACATTGATTAACAATGAACTATTATTATTA 482
  |||||
160 lLysAsnPheSerThrAspAsnGlyAlaIleThrAlaLysThrLeuAs 177
  |||||
483 ACAAGATTACTGTAGGAAATAGCGGAGCCATTCTACCAAGAAATCTTT 532
  |||||
177 eLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer 193
  |||||
533 CTTTGAATAAAGCAGCAGGATCTTCTTTGAAGGAATAAATACGAGC 582
  |||||
194 .....LysLysGlyGlyAlaIleGluThrSerAspAlaLeuThrI 207
  |||||
583 GCACAGGGAATAAAGGTGGGCTATTGTGCTACTGTACTGTAGATAT 632
  |||||
207 eThrGlyAsnGluGlyAlaValSerPheSerAspAsnThrSerSerAsp 224
  |||||

```

```

633 TACAAATTAATACGGCTCCCTACCTCTTCGAAACAATATTCGTGAAGTCG 682
  |||||
224 eGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnSnpA 240
  |||||
683 CAGCTGAGCTATTAATATACACAGGAAACTGACATTTACAGGAATACG 732
  |||||
241 LysValSerPheIleAspAsnLysValThrGlyAlaIleSerSerThrTh 257
  |||||
733 TCTCTGTATTCTTCGAAATAATAGTGTGACA.....GCCAGCCG 770
  |||||
257 rGlyAspMetSerGlyGlyAlaIleCysAlaLysThrSerThrAspT 274
  |||||
771 AGGA.....AATGGAGAGCTCTT.....TCTGGAGATG 799
  |||||
274 hrLysValThrLeuThrGlyAsnGluMetLeuLeuPheSerAsnAsnThr 290
  |||||
800 CCGATGTTACCATATCTGGGAATCAGAGTGAATCTTCTCAGGAACCA 849
  |||||
291 SerThrThrAlaGlyGlyAlaIleTyrrValLysLysLeuGluLeuAlaSe 307
  |||||
307 I.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyGlyT 322
  |||||
900 CGGGGGGGGGGGGTATCTCTCTTTCTACATATATATATCCAGATACA 949
  |||||
322 hrAlaProLysGlyGlyAlaIleAlaIleGluAspSerGlyGluLeuSer 338
  |||||
950 CTGCGAGTAATGCTGGACCATTTCTATCTGCGCAGCAGAGTATGAGT 999
  |||||
339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValIThrSe 355
  |||||
1000 CTTTACGAGAGAGCAGCGGACATTAATCTTCATAGGAAATGCAATGGTGC 1049
  |||||
355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSera 371
  |||||
1050 AACTACACACACAACACTCAAAAGAAATTTATGTAGATATGAGATCTACTG 1099
  |||||
371 lAlysMetThrAlaLeuArgSerIleAlaGlyArgAlaIleTyrrPheTy 387
  |||||
1100 CAATAGACGCAATTTAGTGCATATCTGGCAATATCTGGCATACACTTTTCTAC 1149
  |||||
388 AspProIleThrThrGlySerSerThrThrValThrAspValLeuLysVa 404
  |||||
1150 GATCGAATTACTGCTAATACGGCTGCGGATTCACAGATCTTAATATCTT 1199
  |||||
404 lAsnGluThrProAlaAspSerAlaLeuGluIleTyrrGlyAsnIleLeP 421
  |||||
1200 CAATAGGCTGATGTCAGTATAGTATACAGATTATAGGCGTCAATGTTT 1249
  |||||
421 heThrGlyGlyLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeu 437
  |||||
1250 TTTTGTGTAAGAACTCTCGAAGATGACCAAAAGTTGACAGCAACCTTC 1299
  |||||
438 ThrSerLysLeuLeuGluProValThrLeuSerGlyGlyThrLeuSerie 454
  |||||
1300 ACTTCTACGCTGAAGCAGCTGTAACTTAATCGACAGGAATTTAATACT 1349
  |||||
454 uLysHisGlyValThrLeuGluThrGlnAlaPheThrGlnGluAlaAspS 471
  |||||
1350 TAAACGAGGTGCTACTCTCGATACGAAGGCTTTACTCAGACCGCGGTTT 1399
  |||||
471 eArgLeuGluMetAspValGlyThrThrLeuGluProAla...AspThr 486
  |||||
1400 CCTCTGTATTATATGATCGGCGACACACTTAATAACAAAGTACACAGGAG 1449
  |||||
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLy 503
  |||||
1450 GTCACTTAAACAGGTCTTCCATTCTCTAGACTTTTGGGAGAGGTAA 1499
  |||||
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerG 520
  |||||

```


PT mammals, especially humans -

XX Claim 1; Fig 15-E; 201pp; English.

CC 261502-09 encode Chlamydia pneumoniae polypeptides. The polypeptides
CC are present in the bacterial membrane structure, in the external
CC vicinity of the membrane structure, in the inclusion membrane
CC structure, in the external vicinity of the inclusion membrane
CC and in the cytoplasm of the infected cell. The polypeptides may be
CC used to prevent, treat and detect the presence of Chlamydia infection
CC and/or the presence of Chlamydia in a sample. The polypeptides may
CC also be used to induce an immune response in a mammal. The vaccine
CC vector comprising the polynucleotides is used to induce an immune
CC response in a mammal. Antibodies directed against the polypeptides
CC may also be used therapeutically to treat and/or prevent a Chlamydia
CC infection.

SO Sequence 2957 BP; 818 A; 598 C; 683 G; 858 T; 0 other;

alignment_scores:

Quality: 2006.00 Length: 948
Ratio: 2.810 Gaps: 20
Percent Similarity: 75.316 Percent Identity: 45.570

alignment_block:

US-09-428-122-2 x 261509 ..

Align seg 1/1 to: 261509 from: 1 to: 2957

1 MetyserserPheProlyspheValPheSerThreAlaIlePhePr 17
101 ATGAGATCGTCTTTTCTTCTGTTATTAATCTTCACTTACCTTTCC 150
17 OLeuSerMetIle.....AlaThrGluThrValLeuAspSer 30
151 TCTTATATAGTGTTCGACAGTGTGCGCATTCACATTAGGAGTC 200
30 erAlaSerPheAspGlyAsn...LysAsnGlyAsnPheSerValaGlu 45
201 GTGACAGTTTATATGATGATACAGCACCAAGAAATTTACTCTTAAGCG 250
46 SerGluGluAspAla...GlyThrThrTyLeuPheGlyAsnValTh 61
251 GCACACTTCGATGCTAGTGCAGACCTAATTCGATGGGAGATCTTC 300
61 rIleGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPhe 78
301 GATACCCAA...GCAGGAAACAAACAGCCTTAACCAAGTTGTTT 347
78 snAsnThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeu 94
348 CTACACTCGAGAAATCTTACCTCTTAGGACAGGATTTCTCTTCA 397
95 PheGluThrValaAspAlaGlyThrValaAlaGlyAlaValaAsnSer 111
398 TTTGACAAATTAATTTCTGCTACTGTGAGGTGTTGTTAGCAATAC 447
111 rValaAspLysSerThrThrPheIleGlyPheSerSerLeuSerPhe 128
448 ACCAGCTTGTGGATTACCAATTTCTCAGATTTTCAACTCTTCGATGC 497
128 leAlaSerProGlySerSerIleThrThrGlyLysGlyAlaValaSer 144
498 TTGAGCTCCCTAGG.....ACCACAGTAAGAGGACCATTAAT 538
145 SerThrLysSerLeuThrLysSerValaSerLeuLeuPheSerLeu 161
539 ACCGATGTT...CTGGTGTGAGAGTATAGGAATCTTGAATCTTAATGA 585
161 snAsnPheSerThrAspAsnGlyGlyAlaIleThrAlaLysThrLeuSer 178
586 AAATGCCCTAGTGAATAATGGGGAGGCATCATACGACAGCTTGTCTT 635

178 eutHrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSerLys 194
636 TGACTGGGACTACGGCGTTTGACCGTCTTGGCAATGCTGTCGCA 685
195 LysGlyGlyAlaIleGluThrSerAspAlaLeuThrIleThrGlyAsnG 211
686 CAAGGGGAGCGATCTATGCTCTGTGACTGTGATTTTCAAGAAATGC 735
211 ngIyGluValSerPheSerAspAsnThrSerSerAspSerGlyAlaI 228
736 AGGAATCTTGAGCTTCGAAACAAACAGTCCGACACATCAGAGCGCA 785
228 lePheThrGluAlaSerValThrIleSerAsnAspAlaValaSerPhe 244
786 TCTGCTGCAAGGAACTTGTGATCTCAATTAACCAAAATATCTTTTC 835
245 IleAspAsnLysValThrGlyAlaSerSerSerThrThrGlyAspMet 261
836 GATGGCTGCAGCAACT.....ACAA 858
261 rGlyGlyAlaIleCysAlaTyLys.....ThrSerThrAspThrLys 276
859 TGGCGAGCTATGATTTACAAACGAGGCGAGACCAACCTATCT 908
276 alThrLeuThrGlyAsnGluMetLeuLeuPheSerAsnThrSerThr 292
909 TGACTCTTCAGGAATGAGAGCTGATTTTGAATTAACACAGCAGGA 958
293 ThrAlaGlyGlyAlaIleTyValLysLysLeuGluLeuAlaSerGly 308
959 AATAGTAGAGTGGCATTTATACCAAAATGTTGTATCTCAGAGACG 1008
309 GlyLeuThrLeuPheSerArgAsnSerValaAsnGlyTyThrAlaPro 325
1009 AGGAGAGTGTATTTCTTAACAACAAGCTGCAATGCTACT...CCTA 1055
325 ysgIyGlyAlaIleAlaIleGluAspSerGlyGluLeuSerAla 341
1056 AAGAGGGGCAATTCGATCTCTGAGAGATTTACGATTTCTGCA 1105
342 AspSerGlyAspIleValPheLeuGlyAsnThrValThrSerThrThr 357
1106 GATCTCGCAATATCATTTTCGAGGCAATACT...ACGAGCACTACAG 1152
358ProGlyThrAsnArgSerSerIleAspLeuGlyThrSerAla 372
1153 AAGTCCGCGAGTGTGACGAAATGCTATGATCTTGATCGAATGCA 1202
372 ysnethrAlaLeuArgSerAlaAlaGlyArgAlaIleTyPheTyAsp 388
1203 AATTTTAAATCTCGAGGAGCTCGGGAATTAAGTTATTTTCTATGAT 1252
389 ProIleThrThrGlySerSerThrThrValThrAspValLeuLysVal 405
1253 CCTATACG.....AGCTGAGAGCTACTGATTAACCTCTTTTGA 1293
405 ngIuThrProAlaAspSerAlaLeuGluThrThrGlyAsnIleIlePhe 422
1294 TAAAGCTGACGAGATCTGAAATACATATACGATACATGCTTTCT 1343
422 hrcIyGlyLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeuThr 438
1344 CTGAGAGAAACTCTCAGAAAGAACTTAAGAACTGACATCTGAAG 1393
439 SerLysLeuLeuGluProValThrLeuSerGlyGlyThrLeuSerLeu 455
1394 TCTACATTTACACAGCGCTTAGAGCTTCTGAGGTGCTTAGATGA 1443
455 shiGlyValThrLeuGluThrGluAlaPheThrGluAlaAspSera 472
1444 AGATGAGTACTAGTGTGCAAAATATATACGAGAGGTGCAAGGATCA 1493

```

472 rgleuGluMetAspValAlaIThrThrLeuGlu... ProAlaAspThrSer 487
      ::::: ||||| ||||| ::::: |||||
1494 AAGTCGTAATGAGGAGGACTACTTTTGAAGCAAGCGCTGACGGGGTTC 1543
488 ThrIleAsnLeuValIleAsnIleSerSerIleAspGlyAlaLys 504
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1544 ACTCTCAATGGCCTTAGCCATTATATAGTTCTTAGATGGACAAATTA 1593
504 SALaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerGly 521
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1594 AGCTATCATTTAAGCGGACGGAGCAAGTAGAGTGTGCTTATCAGGGC 1643
521 hrIleThrLeuAsnProThrGlyThrPheTyrGluAsnHisSerLeu 537
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1644 CTATCATCTGTAGATGCTCAGGGAACTATATAGACATCATATATCTC 1693
538 ArgAsnProGlnSerTyrAspIleLeuGluLeuLysAlaSerGlyThrVa 554
      ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
1694 AGTCACACAGCAGCTTCTTCTTATATAGCTTCTGCACAAAGAACAT 1743
554 IThrSerThrAlaValThrProAspProIleMetGlyLysPheHisT 571
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1744 GACTCTACAGATATCCCGATACCCCAATTCATAATCTACGATCAGT 1793
571 YrGlyTyrGlnGlyThrThrGlyProIleValTyr... GlyThr 584
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1794 ATGGGTATCAAGGAGACTGAGT... ATGTGTGGCTGACAGCATCACT 1840
585 GlyAlaSerThrThrAlaThrPheAsnTyrThrLysThrGlyTyrIle 601
      ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
1841 GCAAAACAAAAAATGCTACCTTACTTGAGCTAAACAGATACAGCC 1890
601 CAsnProGluArgIleGlySerLeuValProAsnSerLeuThrPasnAlaP 618
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1891 GAATCCAAAGCTCAGGACCTTTGTCTCAATAGCCGTGGGGTTCCT 1940
618 heIleAspIleSerSerLeuHisTyrLeuMetGluThrAlaAsnGluGly 634
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1941 TTGTGATGTCGCTCCATTCACAGCTCATGACCGGACGACCAAGTTG 1990
635 LeuGlnGlyAspArgAlaPheThrPcysAlaGlyLeuSerAsnPheHis 651
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
1991 TTATCTGCTCAACAATTTGGGTATCAGGATCGCGACTTTTGCA 2040
651 sLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyTyr 668
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2041 TGAAGATCAGAAAGGAACCAACGATGATCTGATCTTACCGGGGT 2090
668 YrValIleGlyLysAsnLeuHisThrCysSerAspLysIleLeuSerAla 684
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2091 ATGCATTAGAGAGAGATTCTTACGGCTTCGAAATTTCTTTAATTT 2140
685 AlaPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAlaLys 701
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2141 GCTTTTGTGAGCTTTTGGCTACACAGACCATCTTGTGGCTAAGAA 2190
701 ngIleGlyThrValTyrGlyThrLeuTyrTrpHis... A 715
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2191 CCATACCATGATATGCAAGGCAATGATGACCGACACTCGAGAGAT 2240
715 snGluThrTyrIleSerLeuProCysLysLeuArgProCysSerLeuSer 731
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2241 CTAAAGCCCTGGCTAAGAT... TGTGCA 2266
732 TyrValProThrGluIleProValLeuPheSerGlyAsnLeuSerTyrTh 748
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2267 GGAATTCGACTCCCTACCTTTGCTTCAATGCTCGTTTGTCTTATGC 2316
748 HisThrAspAsnAspLeuLysThrLysTyrThrThrThrProThrVal 765
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2317 CCATACCGCATATACATGACCAAGATACACTGCTATTCTCTGTTA 2366
765 ysgIySerTyrPcylAsnAspSerPheAlaLeuGluPheGlyLysArgAla 781

```

```

||||| ::::: ||||| ::::: ||||| ::::: |||||
2367 AGGAGAGCTGGGAAATGATGCTTCCTCGTATGATATGTGAGGAGCTATC 2416
782 ProIleCysLeuAsp... GluSerAlaLeuPheGluGlnTyrMetCProPh 797
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2417 CCGGTAGTGTGCTTACAGACGTCGGTCTTGGTGATGATCCACAGCCATT 2466
797 emetLysLeuGlnPheValTyrAlaHisGlnGluGlyPheLysGluGln 814
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2467 TCTAACCTAGATGATCATCTATCCATCAGAAATGATCTTAAAGAAAAG 2516
814 IYThrGluAlaArgGluPheGlySerSerArgLeuValAsnLeuAlaLeu 830
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2517 GCACGAAAGCGCTTCTTCCAAAGTGAAGACCTTCAATCTAGCGGCT 2566
831 ProIleGlyIleArgPheAspLysGluSerAspCysGlnAspAlaThrTy 847
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2567 CCTGATAGGATAAATTTGAAGAAATCTCCGAT... AAGCTACGTA 2610
847 rAsnLeuThrLeuGlyTyrThrValAspLeuValArgSerAsnProAspC 864
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2611 TGATCTCTCCATACCTTACGTTCCCGATGTGATTCGTAATGATCCAGCT 2660
864 ySThrThrThrLeuArgIleSerGlyAspSerTyrPlyThrPheGlyThr 880
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2661 GCACGACACACTTATAGTTCTTCTGGGATCTTCTGACACATGTGSTRCA 2710
881 AsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisPheCysPh 897
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2711 AGCTGTGCTACACAAGCTCTTCTGTACGTCTCGAATCATCATGCCCT 2760
897 eAsnSerAsnPheGluAlaPheSerGlnPheSerPheGluLeuArgGlyS 914
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2761 TGCTTCAACCTTGAAGTTTCAGTCACTTGAAGTCGAGTGGAGGCT 2810
914 eSerArgAsnTyrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
2811 CTCTCTGATGATGATGATCATCTTGTGAGAGAAATTCGGATT 2854
seq_name: /SID6/gcgdata/geneseq/geneseqn/NA1999.DAT.X06821
seq_documentation_block:
ID X06821 standard; DNA: 2757 BP.
XX
AC X06821;
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp9 DNA.
XX
KW Omp9; outer membrane protein 9; surface exposed protein; antigen;
  infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
OS Chlamydia pneumoniae.
XX
PN W09858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
XX
PA (CHR/) CHRISTIANSEN G.
XX
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
  Mydind P;
XX
DR WPI: 1999-105610/09.
XX
DR P-PSDB: W06822.
XX
PT Species-specific test for identifying mammals infected with

```

PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins

PS Claim 6; Page 55-56; 115pp; English.

CC This DNA sequence codes for the novel 96.7 kDa surface exposed
CC protein Omp9 (see W88422) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see X06816-27) encoding
CC Omp4-Omp5 proteins (see W88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.

50 Sequence 2757 BP; 770 A; 564 C; 642 G; 781 T; 0 other;

alignment_scores:

Quality: 2001.00 Length: 953
Ratio: 2.814 Gaps: 21
Percent Similarity: 74.607 Percent Identity: 45.435

alignment_block:

US-09-428-122-2 x X06821 ..

Align seg 1/1 to: X06821 from: 1 to: 2757

```

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
1 ATGAGATGCTCTTTTCTTCTGTTATTAATCTTCATCTAGCCCTTTC 50
17 OleuSerMetIle.....AlaThrGluThrValLeuAspSerS 30
1 TCTCTTAATGAGTGTTCCTGCAGATGCTGCCGATCTCATTAGGAGTC 100
30 eAlaSerPheAspGlyAsn...LysAsnGlyAsnPheSerValArgGlu 45
101 GTGACAGTTAATGATGATACAGACACACAGAAATTACTCTAAAGCG 150
46 SerGlnGluAspAla...GlyThrThrTyrLeuPheLysGlyAsnVal 61
151 GCACATCTGTGATCTAGTGCACGACCATTAATTTCTCATGGGATGCTC 200
61 rLeuGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPhe 78
201 GATAAGCCAA...GCAGGAAACAACAAGAGAGCTTAACACAAAGTTT 247
78 snAsnThrLysGlyAspLeuThrPheThrLysGlyAsnSerLeuLeu 94
248 CTAGCACTGAGGAATCTTACCTTTAGGGAACGATTTCTCTTCAT 297
95 PheGlnThrValAspAlaGlyThrValAlaGlyAlaValAlaSerSe 111
298 TTGACATAATATTATTGCTACTGTCGAGGTGTTGTTGTTAGCAATAC 347
111 rValValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPhe 128
348 AGCAGCTTCTGGATATGAAATTCAGAGATTTTCAACCTTCGAGATGC 397
```

```

128 leAlaSerProGlySerSerIleThrThrGlyGlyAlaValSerCys 144
398 TTGCAGCTCTAGC.....ACCACAGTAAAGAGCATTAAAT 438
145 SerThrGlySerLeu.....SerLeuThrLysAsnValSe 156
439 ACCGATGCTCTGGTGTGTTGAGAGATATGAGGAATCTTACCAAAAT 483
156 rLeuLeuPheSerLysAsnPheSerThrAspAsnGlyAlaIleThr 173
484 .....GAAATGCCCTACTGAAATATGGGAGGACCATCATATA 520
173 lAluThrIleSerLeuThrGlyThrThrMetSerAlaLeuPheSerGlu 189
521 CGAAGACTTGTCTTTGACTGGGAGTACGGGTTTGTAGCGCTTCCTGGC 570
190 AsnThrSerSerLysGlyGlyAlaIleGluThrSerAspAlaLeuTh 206
571 AATAGCTCTGCGACACAGGGAGCGATCTATGCTTGTGACTCTGT 620
206 rIleThrGlyAsnGlnGlyValSerPheSerAspAsnThrSerSera 223
621 GATTTCTGAGAAATGCAAGAAATCTTGAGCTTCGGAACAACAGTGCACA 670
223 spSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsn 239
671 CATCAGAGGCGGATCTGCTGAAGGAACTTGATGATCTCCAAATAC 720
240 AlAluValSerPheIleAspAsnLysValThrGlyAlaSerSerThr 256
721 CAAATATCTTTTTCATGCTGCTCAAGCACT..... 753
256 rThrGlyAspMetSerGlyAlaIleCysAlaTyrLys.....Ths 271
754 .....ACAAATGCGGACCTTATGATTGATTAACAAGACAGGCGCA 793
271 eThrAspThrLysValThrLeuThrGlyAsnGluMetLeuLeuPheSer 287
794 ACCCAGACCCATCTGACTCTTTCAGAAATGAGAGCCTGCAATTTCTG 843
288 AsnAsnThrSerThrThrAlaGlyAlaIleIleValLysLysLeuG 304
844 AATACACAGCAGCAAGAAATAGTGAGGTGCGATTATACCAAAAATGTGT 893
304 uLeuAlaSerGly...GlyLeuThrLeuPheSerAsnSerValAsnG 320
894 GTTATCTCAGACGAGAGAGGTATTATTCTTAACCAAAAGCTGCGA 943
320 lGlyThrAlaProLysGlyAlaIleAlaIleGluAspSerGlyGlu 336
944 ATGCTACT...CCTAAGAGAGGGCAATTGCGATTCTAGATTCTGAGAG 990
337 LeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThr 353
991 ATTAGATTCTGCGAGATCTCGCAATATCATATTTCGAGGCAATAC 1038
353 lThrSerThrThr.....ProGlyThrAsnArgSerSerIleAspL 367
1039 .ACGACCATCTACAGAGAGCTCGCGAGTGTGACCAAGAATGCTATAGATC 1087
367 euGlyThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAla 383
1088 TTGCATCGCAATCAAAATTTTAATCTCCGAGCGACCTCGGGAAATAAA 1137
384 lIleTyrPheTyrAspProIleThrThrGlySerSerThrThrValThr 400
1138 GTTATTCTTCTATGATCCATACG.....AGCTCAGAGAGCTACGTA 1178
400 pValLeuValAsnGluThrProAlaAspSerAlaLeuIleThrThrG 417
1179 TTAGCTCTCTTGAATTAAGCTGACGACGAGATCTGGAATATCTTATGAG 1228
417 lYAsnIleIlePheThrGlyGluLysLeuSerGluThrGluAlaAlaAsp 433
```



```

1229 GCTACATGCTTTCTCTGGAGAACTCTCAGAGAGCACTTAAGAA 1278
1230 SerLysAsnLeuThrSerLysLeuGlnProValThrLeuSerGly 450
1279 CCTGACATCTCAAGTCTCATTTACACAGGCTGTAGACCTTGCACG 1328
450 yThrLeuSerLysHisGlyValThrLeuGlnThrGlnAlaPheThr 467
1329 TGCCTTAGATTGAAAGATGAGTACGTAGTCAATACTATAACGC 1378
467 InGlnAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuGlu 482
1379 AGGTGAGGAGCGAAAGCTGTTATGATGAGGAGGACTCTTTAGGCA 1428
483 ProAlaAspThrSerThrIleAsnAsnLeuValIleAsnIleSerSer 499
1429 AGCGCTGAGGGGCTCATCTCATAGGCTTATATATATATGATTCMT 1478
499 eAspGlyAlaLysLysAlaLysIleGluThrLysAlaThrSerLysAsn 516
1479 AGATGGGACAAATTAAGCTATCAAGCGAGCGACGACAACTAAGATG 1528
516 euThrLeuSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyr 532
1529 TTGCCTTATCAGGGCTTATCATGCTTGTAGATGCTCAGGGACATAT 1578
533 GluAsnHisSerLeuArgAsnProGlnSerTyrAspIleLeuGluLeu 549
1579 GAGCATCATATCTCAGTCACAGCAGGCTTCTTATATAGACTTTC 1628
549 sAlaSerGlyThrValThrSerThrAlaValThrProAspProIleMet 566
1629 TGCACAGAGAAAGATGACTACTACAGATATCCCGATACCCCAATTC 1678
566 yGluLysPheHisTyrGlyTyrGlnGlyThrTrpGlyProIleValTrp 582
1679 ATACTACCAATCTAGTGGTATCAAGAACT...GGAATATTTGTTG 1725
583 GlyThrGlyAlaSerThrThrAlaThrPheAsnTrpThr 596
1726 GTGACGATGCAACTGCACAAAACAAATGCTTACCTTAACCTGAC 1775
596 sThrGlyThrIleProAsnProGluArgIleGlySerLeuValProAsn 613
1776 AACAGATATCAAGCCGAATCCAGAAAGTCAGGACCTTGTGCTTAATA 1825
613 eThrLeuTrpAsnAlaPheIleAspIleSerSerLeuHisTyrLeuMetGlu 629
1826 GCCTGTGGGGTCTTTGTCATGTCGCTCATCAGAGCCTCATGAGC 1875
630 ThrAlaAsnGluGlyLeuGlnGlyAspArgAlaPheTrpCysAlaGly 646
1876 CGGACGACAAAGTTCGTATCTTCGTCACAAATTTGTGGTTCAGGAAT 1925
646 uSerAsnPhePheHisLysAspSerThrLysThrArgArgLysPheArg 663
1926 CCGGACCTTTTGCATGAGATCAGAAAGAAACCAACCTATATTCGTC 1975
663 sLeuSerGlyLysValIleGlyGlyAsnLeuHisThrCysSerAsp 679
1976 ATTCTAGCGCGGTATGCATAGAGAGAGATTCCTCAGCGCTCTGAA 2025
680 LysIleLeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAsp 696
2026 AATTTCTTAATTTGCTTTGTAGCTTTTGGCTACGACGAAAGACCA 2075
696 rPheValAlaLysAsnGlnGlyThrValTyrGlyLysThrLeuTyrTyr 713
2076 TCTTGTGCTAGAACCATACCATGTATATCAGGCGCAATGAGATTACC 2125
713 InHis.....AsnGluThrTyrIleSerLeuProCysLysLeuArg 726

```

```

2126 GACACCTCGGAGAGCTAAGACCTCGCTAAGATT..... 2160
727 ProCysSerLeuSerTyrValProThrGluIleProValIlePheSer 743
2161 .....TGTCAAGAAATTTGACTCCCTACTCTTGTGCTTCAATGC 2201
743 yAsnLeuSerTyrThrHisThrAspAsnAspLeuLysThrLysTyrThr 760
2202 TCGGTTTCTTATGGCCCTFACCGCAATATACATGACCAAAAGTATAC 2251
760 hTyrProThrValLysGlySerTrpGlyAsnAspSerPheAlaLeuGlu 776
2252 GCTATCTCTCTTAAAGGAACTGGGAAATGATGCTTGTGATTAAGA 2301
777 PheGlyIleArgAlaProIleCysLeuAsp...GluSerAlaLeuPhe 792
2302 TGTGAGAGAGATATCCCGGTAGTTCCTCAGAGAGCTGCTTGGGTGA 2351
792 uGlnTyrMetProPheMetLysLeuGlnPheValTyrAlaHisGlnGlu 809
2352 TACCCACAGCCATTCTTAAACCTAGAGATGATCTATGACATCAGATG 2401
809 yPheLysGluGlnGlyThrGluAlaArgGluPheGlySerSerArgLeu 825
2402 ACTTTAAGGAAACGCGACAGAGCCGTTCTTCCAAAGTGAAGACTC 2451
826 ValAsnLeuAlaLeuProIleGlyIleArgPheAspLysGluSerAsp 842
2452 TTCAATCTAGCGGTTCCGTAGGATATAAATTGAGAAATCTCCGAT 2499
842 sGlnAspAlaThrTyrAsnLeuThrLeuGlyTyrThrValAspLeuVal 859
2500 ...AAGTCTAGTATGATCTCTCATAGCTTACCTCCGATGTGATTC 2545
859 rGSerAsnProAspCysThrThrThrLeuArgIleSerGlyAspSerTrp 875
2546 GTATGATCCAGCGCTCGACGAACTGTATAGTTCGTGGGATTCCTGG 2595
876 LysThrPheGlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAla 892
2596 TCGACATGTGTACAGCTTGTCTAGACAAACCTCTTGTGACGTGCGG 2645
892 yAsnHisPheCysPheAsnSerAsnPheGluAlaPheSerGlnPheSer 909
2646 AATTCATCATCCTTGTCTTAAACTTTGAGAGTTTTCAGTAGTTTGA 2695
909 heGluLeuArgGlySerSerArgAsnTyrAsnValAspLeuGlyAla 925
2696 TCGAGTTTCAGAGTTCTTCTGCTAGCTATGATTCATGAGTTCGAG 2745
926 TyrGlnPhe 928
2746 TTCGGATT 2754

```

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT.X06828

seq_documentation_block:
ID X06828 standard: DNA: 3000 BP.
XX X06828;
XX
DT 26-APR-1999 (first entry)
XX
XX
DE Chlamydia pneumoniae surface exposed protein Omp5 DNA.
XX
XX Omp5; outer membrane protein 5; surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
XX Chlamydia pneumoniae.
XX
FH key location/Qualifiers
FT CDS 259..3000
FT /*tag= a

XX W09858953-A2.
 PN
 XX 30-DEC-1998.
 PD
 XX 19-JUN-1998; 98WO-DK00266.
 PF
 XX 23-JUN-1997; 97DK-0000744.
 PR
 XX (BIRK/) BIRKELUND S.
 PA (CHRIL/) CHRISTIANSEN G.
 XX
 P1 Birkelund S, Christiansen G, Knudsen K, Madsen A;
 P1 Mygind P;
 XX WPI: 1999-105610/09.
 DR P-PSDB: W88429.
 DR
 XX
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins

PS Disclosure: Page 73-77; 115pp; English.

XX This DNA sequence encodes the novel surface exposed protein Omp5
 CC (see W88429) of Chlamydia pneumoniae, a human respiratory pathogen.
 CC It is described as a subsequence of a claimed nucleic acid fragment
 CC (see X06817) encoding Omp5 (see W88418). The invention provides a
 CC new species specific test for identifying mammals (including
 CC humans) infected with C. pneumoniae. The test comprises detecting
 CC antibodies specific for surface exposed proteins Omp4-Omp5 (see
 CC W88417-28) or detecting nucleic acid fragments encoding them (see
 CC X06816-27), especially by PCR. The proteins are also used in the
 CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids
 CC and proteins can also be used in the immunization of mammals, the
 CC nucleic acids being particularly useful as DNA vaccines for
 CC effecting in vivo expression of antigens. The vaccines may also
 CC prevent atherosclerosis and bronchial asthma, which are possibly
 CC associated with C. pneumoniae.

SQ Sequence 3000 BP; 867 A; 597 C; 658 G; 878 T; 0 other;

alignment_scores:
 Quality: 2000.00 Length: 936
 Ratio: 2.801 Gaps: 20
 Percent Similarity: 76.282 Percent Identity: 46.261

alignment_block:
 US-09-428-122-2 x X06828 ..

Align seg 1/1 to: X06828 from: 1 to: 3000

1 MetLysSerSerPheProLysPheValPhe...serThrPheAlaIlePh 16
 259 ATGAATTCGCATTTTCCTGCTAGTCTCTCTCGACATTTGGCATGTTT 308
 16 e.....ProLeuSerMetIleAlaThrGluThrValLeuAspSer 30
 309 TACTAGTTGTTCCACTGTTTTCCTGCAACGCTGAATAATAGGCCCT 358
 30 eAlaSerPheAspGlyAsnLysAsn...GlyAsnPheSerValArgLys 45
 359 CTGATACCTTTGACGAGAGTACTACACAGCAGCTTACTCTCTAATAAT 408
 46 SerGlnGluAspAlaGlyThrThrTyrLeuPheLysGlyAsnValThrLe 62
 409 ACG.....ACTACTGGAATAGACTATACCTGACAGAGATATACTCT 452
 62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsn 79
 453 GCAAAACCTT...GGGATTCGCGACGCTTACAGAAAGGTTGTTTCTG 499

79 snThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPhe 95
 500 ACACACGCAATCTTAACTTTGCGCGTAGGGGTACTCTCTCTTT 549
 96 GlnThrValAspAlaGlyThrValAlaGlyAlaValAsnSerSerVal 112
 550 TTAATATTAACTCT...AGTCTGAAGCGCACGACTT...TCTGTAC 593
 112 lValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleA 129
 594 AACTGATTAATAATCTGCGCTACACAGATTTTCAGCTTACTTCTTAG 643
 129 lAserProGlySerSerIleThrThr.....GlyLysGlyAlaValSer 143
 644 CGGCCCATCATCGTAAATCAACCCCTCAGAAAGGTCAGTTAA 693
 144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPhe 160
 694 TGTGGA...GGGATCTTACATTGATACAAATGACATTTATTATTA 740
 741 ACAAGATTACTGTGAGCAAAATGCGGAGCCATTCTACCAAGATCTT 790
 160 rLysAsnPheSerThrAspAsnGlyAlaIleThrAlaLysThrLeu 177
 177 erLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSer 193
 791 CTTTGAAATACAGACGCGATTCGATTTCTTGAAGGATTAATAGTAC 840
 194LysLysGlyAlaIleGlnThrSerAspAlaLeuThrI 207
 841 GCAACAGGAAATAGGCGGTATTGCTACGACTGACTATATAT 890
 207 eThrGlyAsnGlnGlyGluValSerPheSerAspAsnThrSerAsp 224
 891 TACAAATTAATACGCGCTCCCTCTTCGAAACATATGCTGAGCTG 940
 224 erGlyAlaIlePheThrGluAlaSerValThrIleSerAsnAla 240
 941 CAGGTGAGCTATTAATACACAGAACTGATCAATTCAGAGGATACG 990
 241 LysValSerPheIleAspAsnLysValThrGlyAlaSerSerThr 257
 991 TCTCTGTATTCTGAAATAGTGTGACA.....GCGACGCC 1028
 257 rGlyAspMetSerGlyAlaIleLysAlaIleLysAlaIleLysThr 274
 1029 AGGA.....AATGAGAGCTCTT.....TCTGAGATG 1057
 274 hrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsnThr 290
 1058 CCGATGTACCATATCTGGAATCAGAGTCACTTCTCAGAAACCA 1107
 291 serThrThrAlaGlyAlaIleLysValLysLysLeuGluLeuAla 307
 1108 GCTGTAGCTAATGCGGAGCCATTATGCTAAGAGCTTACAGCTGCTC 1157
 307 r.....GlyGlyLeuThrLeuPheSerArgAsnSerValAsnGly 322
 1158 CGGGGGGGGGGGGATCTCTCTTTTCTACATATATATCCAAAGTACCA 1207
 322 hrAlaProLysGlyAlaIleAlaIleGluAspSerGlyGluLeuSer 338
 1208 CTGACAGTATGCTGAGCATTCTATATCTGAGCAGCTGAGAGTGT 1257
 339 LeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValThr 355
 1258 CTTTAGCAGAGACAGGAGCACTTCAATGGAATGCAATGTTGTC 1307
 355 rThrThrPro...GlyThrAsnArgSerSerIleAspLeuGlyThrSer 371
 1308 AACTACACCACTACAAAGAAATTTATTTGACATAGACTCTACTG 1357

```
371 lalysethrAlaleuArgSerAlaIaGlyArgAlaIleTyrPheTyr 387
|||||
1358 CAAGAGTACAGAAATTACGTGCAATATCTGGCATAGCATCTTTCTTAC 1407
388 ASPProIleThrThrGlySerSerThrThrValThrAspValLeuIleuSva 404
|||||
1408 GATCCGATTACTACTATACGGCTGCGGATTCTTACATACTTTAAATCT 1457
404 lAsnGluThrProAlaAspSerAlaLeuGlnTyrThrGlyAsnIleIleP 421
|||||
1458 CAATAAGCGTGATGCAAGGTAATAGTACAGATTATAGCGGTGCTTCTT 1507
421 heThrGlyGlyLeuSerGlnThrGlnAlaAlaAspSerLysAsnLeu 437
|||||
1508 TTTCTGTGTAAGGCTCTGGAAGATGAACAAAGTTGCAGACAACCTC 1557
438 ThrSerLysLeuGlnProValThrLeuSerGlyThrLeuSerLeu 454
|||||
1558 ACTTCTACGCTGAAGCAGCTTACTACTTACGACAGAAATTTAGTACT 1607
454 uLysHISGlyValThrLeuGlnThrGlnAlaPheThrGlnAlaAspS 471
|||||
1608 TAAACGTGGTGCTACTCTCGATACGAAGGCTTACTACAGACCGCGGT 1657
471 eArgLeuGluMetAspValGlyThrThrLeuGluProAla...AspThr 486
|||||
1658 CCTCTGTATTATGATGCGGACACAGCTTAAAGCAAGTACAGAGAG 1707
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAla 503
|||||
1708 GTACATTACAGCTTCTTCCATCTCTGACTCTTACGCGAGGTAA 1757
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSer 520
|||||
1758 GAAAGTTGTAATGCTGCTCTGACCAAGTAAATAATGAGCCCTTAGTG 1807
520 lYThrIleThrLeuLeuAspProThrGlyThrPheThrGlyAsnHISer 536
|||||
1808 GTCCGATTCTCTTTTGGATACCAAGGAAATGCTTGAATAATCAGAC 1857
537 LeuAlaAsnProGlnSerLysAspIleLeuGluLeuLysAlaSerGlyTh 553
|||||
1858 TTAGGAAATCAAGACTTTTCTATTGTGACACTCTCTCTGCGGTAC 1907
553 rValThrSerThrAlaValThrProAspProIleMetGlyGluLysPhe 570
|||||
1908 TGCACACACTACAGATGTTCCACGCGTTCTCTACAGTACACTCTTAC 1957
570 lSTyrGlyTyrGlnGlyThrTyrGlyProIleValTyrPhe... 582
|||||
1958 ACTATGGGTATCAAGTACTTGGGA...ATGACTTGGGTGATGATAC 2004
583 ...GlyThrGlyAlaSerThrThrAlaThrPheAsnTrpThrLysThrG 598
|||||
2005 GCACAGACTCTCAAGACTTAAGACAGCAGCATTTAGCTTGACCAATAC 2054
598 lYThrIleProAsnProGlnArgIleGlySerLeuValProAsnSerLeu 615
|||||
2055 CTACCTTCCGAATCCTGAGCGGTCAAGAGCACTTATGTTCCATAATAC 2104
615 rPAsnAlaPheIleAspIleSerSerLeuHISThrLeuMetGluThrAla 631
|||||
2105 GGGGACTTTTTCAGACATCCAGACGATTCAGGTGTCATAGAGAGAGT 2154
632 AsnGluGlyLeuGlnGlyAspArgAlaPheThrGlyAlaGlyLeuSer 648
|||||
2155 GCTTTGACTCTTTGTCAGATCAGAGCTTCTGGCGGAGGATCGGCAA 2204
648 nPhePheHISLysAspSerThrLysThrArgArgGlyPheArgHISLeu 665
|||||
2205 TTTCTTAGATTAAGATAGAAAGGGAACCAATATACCTCATTAAT 2254
665 eArgLysGlyTyrValIleGlyLysAsnLeuHISThrLysSerAspLysIle 681
```

```
|||||
2255 CTGGTGGAATATGCTATACGGAGGCGACGCCAACTTGTCGAAACTTA 2304
682 LeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArgAspThrPheVa 698
|||||
2305 ATTAGCTTTGCCCTTTTGGCACTCTTTGGTAGCGATTAAGATTCTTAGT 2354
698 lAlaLysAsnGlnGlyThrValTyrGlyGlyThrLeuTyrTyrGlnHISa 715
|||||
2355 CGCTAAATAATCATATGATACCTATGACAGAGCGCTTCATATACCAAC 2403
715 sncGluThrTyrIleSerLeuProCysLysLeuArgProCysSerLeuSer 731
|||||
2404 .....ATTACGAATGATGAGGTGATACGTTGCTCTTAGAT 2442
732 TyrValPro.....ThrGluIleProValLeuPheSerGlyAsnLe 745
|||||
2443 AACTTCCTGCGCTCTGGAGTCAATAACCCCTGTTTGAAGGCGAGCT 2492
745 uSerTyrThrHISThrAspAsnAspLeuLysThrLysTyrThrTyrP 762
|||||
2493 CGCTTTATGCCACGCTCAGTAATGATCTGAAGACAAAGTATACGCTATC 2542
762 roThrValLysGlySerTyrGlyAsnAspSerPheAlaLeuGluPheGly 778
|||||
2543 CTGAGGTGAAGAGTCTTTGGGGAATATGCTTTTAACATGATGTTGGCA 2592
779 GlyArgAlaProIleCysLeuAspGluSerAlaLeuPheGluGlnTyr 795
|||||
2593 GCTTCTTCATCTCTTATCTGATACCTGATACCTGATGTTGATACCTAT 2642
795 tProPheMetLysLeuGlnPheValTyrAlaHISGlnGluGlyPheLys 812
|||||
2643 TCCATCATCAAACTGAATCTGACCTATATACGTCAGACAGCTTCTCGG 2692
812 lGlnGlyThrGlnAlaArgGluPheGlySerSerArgLeuValAsnLeu 828
|||||
2693 AGAAGAGTACAGAGAGAGATCTTTGATGACAGCAACCTTCATTTTA 2742
829 AlaLeuProIleGlyLysArgPheAspLysGluSerAspCysGlnAspAl 845
|||||
2743 TCTTTTCCATATAGGGGTGAAGTTGAGAAGTCTCTGATTGTAATGACTT 2792
845 aThrTyrAsnLeuThrLeuGlyTyrThrValAspLeuValArgSerAsp 862
|||||
2793 TTCTTATGATCTGACTTATCTCATGTCTCTGATCTTATCCGCAATGATC 2842
862 roAspCysThrThrThrLeuArgIleSerGlyAspSerTyrPheThrPhe 878
|||||
2843 CCAATGCACTACAGCATTTGTAATCAGCGGAGCCCTTGGGAACTTAT 2892
879 GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHISph 895
|||||
2893 GCCAATTAATCTTACAGCAGACAGCGCTTGCAAGTGGCGGAGGAGTCT 2942
895 eCysPheAsnSerAspPheGluAlaPheSerGlnPheSerPheGluLeu 912
|||||
2943 CGCCTTCTCTCCATATGTTGAAGTCTCGGCCAGATTGTCTTTGAAGTTC 2992
912 rGlyLysr 914
|||||
2993 GTGGATCC 3000
```

seq_name: /Sids6/gcdata/geneseq/geneseq/NA1999.DAT:X06822

seq_documentation_block:
ID X06822 standard; DNA: 2787 BP.

XX X06822:
AC
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp10 DNA.

XX Omp10: outer membrane protein 10: surface exposed protein: antigen;
 KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
 XX Chlamydia pneumoniae.
 OS
 XX WO9858953-A2.
 PN
 XX 30-DEC-1998.
 PD
 XX 19-JUN-1998; 98WO-DK00266.
 PE
 XX 23-JUN-1997; 97DK-0000744.
 PR
 XX (BIRK/) BIRKELUND S.
 PA (CHR1/) CHRISTIANSEN G.
 XX
 XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mygind P;
 XX WPI: 1999-105610/09.
 DR P-PSDB: W88423.
 XX
 XX Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 PS
 XX Claim 6; Page 59; 115pp: English.
 CC This DNA sequence codes for the novel 98.4 kDa surface exposed
 CC protein Omp10 (see W88423) of the human respiratory pathogen
 CC Chlamydia pneumoniae. By generating antibodies against C.
 CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
 CC was obtained which reacted with outer membrane proteins. The
 CC antibody was used to identify the genes (see X06816-27) encoding
 CC Omp4-omp5 proteins (see W88417-28) in an expression library of
 CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
 CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
 CC the other, and encode polypeptides of about 89.6-100.3 kDa and
 CC about 56.1 kDa. The invention provides a new species specific test
 CC for identifying mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-omp5 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used
 CC in the diagnosis of C. pneumoniae infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting in vivo expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,
 CC which are possibly associated with C. pneumoniae.
 XX
 XX Sequence 2787 BP: 815 A; 689 C; 535 G; 748 T; 0 other;
 SQ
 alignment_scores:
 Quality: 1986.00 Length: 950
 Ratio: 2.770 Gaps: 19
 Percent similarity: 75.474 Percent identity: 45.263
 alignment_block:
 US-09-428-122-2 x X06822 ..
 Align seg 1/1 to: X06822 from: 1 to: 2787
 1 MetlysserSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
 1 ATGAATCTCTCTCATCGTTTATACTCTGTCATCTTTAGCACTTC 50
 17 OleuSerNet.....IleAlaThrGluThrValLeuA 28
 51 CTTGCTACTAATATTCTCTGCGTTGCTGCTGTGTAAGAAATCAATCTAG 100

28 sPserSerAlaSerPheAspGlyAsnLysAsnGlyAsnPheSerValArg 44
 101 GACCTACCAATAGCTTCTCTGGA.....CCAGAACCTACACTCTCCCA 144
 45 GluSerGlnGluAspAla...GlyThrThrTyrLeuPheLysGlyAsnVa 60
 145 GCCCAACACAAATAGCATGAGACTATCTATATCTAACAGGGAGTCT 194
 60 lThrLeuGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysP 77
 195 CTCACATCACCAAT...GCAGATCTCCGACAGCTCTAACCCCTTCCTCT 241
 77 heAsnAsnThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeu 93
 242 TTAAGCAAACTCTCTGGAATCTTTCTTCCAAAGCCAGCGCTACCAATTT 291
 94 LeuPheGlnThrValAspAlaGlyThrValAlaGlyAlaValAsnSe 110
 292 CTCCTACAAATATCGATCGGGA.....GCAGACTGTACTTACCCA 335
 110 rSerValValAspLysSerThrThrPheIleGlyPheSerSerLeuSerP 127
 336 TACACCTCCAAATAGCTTCTCTCTTCAGGATTCCTATTGTCTC 385
 127 heIleAlaSerProGlySerSerIleThrThrGlyLysGlyAlaValSer 143
 386 TAATACAAAC.....ACGAATGCTACACAGACAGAGCCATCAAG 429
 144 CysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSe 160
 430 ...TCCACAGAGCGCTGTGTTTATTACGTGCACTAATAGTTGCTACTTTGG 476
 160 rLysAsnPheSerThrAspAsnGlyGlyAlaIleThrAlaLysThrLys 177
 477 CCAAACTTTCTTAATGACATGAGCGCCCTCCAAAGCGCTCTATCA 526
 177 erLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer 193
 527 GTCATATCG...CTAAACCCCACTAAGCTTTGCCAAACCAAAAGCAACG 573
 194 LysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAs 210
 574 CAABAAGGGGGTGCCTTATTCACGCGAGGAGTTCAAATTAACAATAC 623
 210 nGlnGlyGluValSerPheSerAspAsnThrSerSerAspSerGlyAla 227
 624 GTTAACACACATCATTTCTTGAAATACCGCGCCCAACAATGGCGGAG 673
 227 lAlaIlePheThrGluAlaSerValThrIleSerAsnAsnAlaLysValSer 243
 674 CCATTTCACGGAAGCTAGAGTTTATTAGACACAAAGCAATATAGC 723
 244 PheIleAspAsnLysValThrGlyAlaSerSerSerThrThrGlyAspMe 260
 724 TTTATAAACATAGTGTGACCGCAACTCAGCT..... 756
 260 tSerGlyGlyAlaIleCysAlaTyrLysThrSerThrAspThrLysVal. 276
 757 .ACAGGGGAGCCATTACTGTAGTATACATCAGCCCAACCAACGACTCT 805
 277 ..ThrLeuThrGlyAsnGluMetLeuLeuPheSerAsnAsnThrSerThr 292
 806 TAACCTATATCAGACACAGGGAAGTGAACCTTATAGGAATATACGCAAT 855
 293 ThrAlaGlyGlyAlaIleTyrValLysLysLeuGluLeuAlaSerGly 309
 856 ACTAGTGTGGGGGATTTACTGACAAATCTAGTCTTTCTTCTTGAGG 905
 309 yLeuThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaProLysG 326
 906 ACCTACGCTTTTAAACAACACTGTGCTATAGTACTGACGACCTCCCTTAG 955
 326 LylGlyAlaIleAlaIleGluAspSerGlyGluLeuSerLeuSerAlaAsp 342

```

956 GAGGAGCAATTCGATTCGACTCTGGATCTTGATGCTTTGGCTCTT 1005
343 SerGlyAspIleValPheLeuGlyAsnThrVal.....ThSerTh 356
1006 GGTGGAGCATCACCATTGGAAGAAACACAGTACGAAAGAGCTTCTTC 1055
356 rThProGlyThrAsnArgSerSerIleAspLeuGly...ThSerAlaL 372
1056 GAGTCAGACCACTACAGAAATTCATTACATCGAAACACCAATGCTA 1105
372 ySmetThrAlaLeuArgSerIleAlaGlyArgAlaIleTyrPheTyrSp 388
1106 AGATTGTACAGCTCGAGCCTCTCAGGCAATACTATCTACTCTATGAT 1155
389 ProIleThrThrGlySerSerThrThrValThrAspValLeuLysValas 405
1156 CCTATTAACAATAACCATGCTGACACTCTCGAGATGCTTAACCTTAA 1205
405 nGluThrProAlaAspSerAlaLeuGlnTyrThrGlyAsnIleIlePheT 422
1206 TGGTCTTGACCTTGCAGGGAAATCCTGCATATCAAGAACCATCGATT 1255
422 hrGlyGlyLysLeuSerGlnThrGlnAlaAlaAspSerLysAsnLeuThr 438
1256 CTGGAGAACACTCTCGAGACGAGACGTCAGAACGTCGATATCTCA 1305
439 SerLysLeuLeuGlnProValThrLeuSerGlyGlyThrLeuSerLeu 455
1306 TCTCAATTAATTCAGCAACTCTTAACCTCTCGGGAGGCAACTCTCTT 1355
455 sHisGlyValThrLeuGlnThrGlnAlaPheThrGlnGlnAlaAspSer 472
1356 ATCAAGAGTCACTGATGCTGCTAAGTCTTTTGCATCTCCGGCTCTA 1405
472 rGleuGlnMetAspValGlyThrThrLeuGlnProAlaAspThrSerThr 488
1406 CCTCTCTGATGATGACGAGGACCATTAAGAACCCGTCGATGGATCCT 1455
489 IleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLysLysAl 505
1456 ATCAATAATCTTGTCTCAATGATGATCTTAAGACAGACCAAGAAAGC 1505
505 aLysIleGlnThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrI 522
1506 TACGCTAAACACACACAGCAGTACAGACGTCATTATCTGGATGTC 1555
522 IeThrIleLeuAspProThrGlyThrPheTyrGluAsnHisSerLeuArg 538
1556 TCTCTCTGATGATCTCTGGAATGTCGAAAGATGTCCTCTGGAAT 1605
539 AsnProGlnSerTyrAspIleLeuGlnLeuLysAlaSer.....GlyTh 553
1606 AACCTCTAGCTTTTCTCTCTCACTCTACTCTGTCAGACCCCGGCA 1655
553 rValThrSerThrAlaValThrProAspProIleMetGlyGluLysPheH 570
1656 TATTCACATCACAAGACTAGCTGCGATCCCTAGAAAAAATCTTAATCC 1705
570 IeTyrGlyTyrGlnGlyThrTyrPheGlyProIleValTyrGlyThrGlyAla 586
1706 ATGGGGATACCAAGGAAATGGGCA...TATCTTGGCAGAGGATCT 1752
587 SerThr.....ThrAlaThrPheAsnThrPheThrLysThrGlyTyrI 600
1753 GCGACTAAATCAAGACGACTCTTACCTGACAAAAACAGGATCAAA 1802
600 eProAsnProGlnArgIleGlySerLeuValProAsnSerLeuTyrAsnA 617
1803 TCCGAATCTCAGCGTCTGAGACTTAGTGTGCAACGCTATAGGGAT 1852
617 IaPheIleAspIleSerLeuHisTyrLleuMetGlnThrAlaAsnGln 633
:::
1853 CCTTTGATATGTCGCTCCATACACAGCTTGTAGCCTAAAGTACGC 1902
634 GlyLeuGlnIleAspArgAlaPheThrCysAlaGlyLeuSerAsnPheP 650
1903 CAATCTCAGAAATCGCGGATGCTGTTGAAGGATCTCGAACTTCT 1952
650 eHisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGlyG 667
1953 CCATTAAGATGACGAGAAATTAAGGTTTCCACATATAGTGCAG 2002
667 LyTyrValIleGlyLysLeuHisThrCysSerAspLysIleLeuSer 683
2003 GTTATGTTGAGAGGACACTCAACATTAGCTTGTGATTAATCTTATACT 2052
684 AlaAlaPheCysGlnLeuPheGlyArgAspArgAspTyrPheValAla 700
2053 GCACCTCTCTCCATATATTCGGGAAGATAGAGATCATTATTAATTA 2102
700 sAsnGlnGlyThrValTyrGlyGlyThrLeuTyrTyrGlnHisAsnGln 717
2103 AAATAGACTTCTGCTATGAGCTTCTCTCACTCCACACTCAGACTAG 2152
717 hrTyrIleSerLeuProCysLysLeuArgProCysSerLeuSerTyrVal 733
2153 CCTTGCTCTCT.....CCAGCTGTATACGCTACTT 2184
734 Pro.....ThrGlnIleProValLeuPheSerGlyAsnLeuSerTyr 747
2185 CCTGATCTGAAGTGAAGTCAAGCCTCTCTTGTGATCTCAGATCACCTA 2234
747 rThrHisThrAspAsnLeuLysThrLysTyrThrThrTyrProThrV 764
2235 TATCTATAGTAAATCTATGAAGACCTATTACACCAAGCACAAGC 2284
764 aLysGlySerTyrPheLysAsnAspSerPheAlaLeuGlnPheGlyArg 780
2285 GAGAGACTCTGCTGATATATGACGCTTCGCTCGAATCTCGAGCTCC 2334
781 AlaPro...IleCysLeuAspGlnSerAlaLeuPheGlnGlnTyrMetPr 796
2335 CTACACACACACTGTTTAAGCAGAGAGGCTCTTCACCGATATTTTCC 2384
796 oPheMetLysLeuGlnPheValTyrAlaHisGlnGlyPheLysGlnG 813
2385 TTTATTAAGATAGAACCTTCGTACATACACCAAGATAGCTTCAAGAAC 2434
813 InGlyThrGln...AlaArgLysPheGlySerSerArgLeuValAsnLeu 828
2435 GTATATCTACTCTGTCGATCTTTCGATGCGGATTTAATTAATTAAGCTC 2484
829 AlaLeuProIleGlyIleArgPheAspLysGlnSerAspCysGlnAspAl 845
2485 TCTGTGCTATGGAATTAACCTTCGAGAGATTCGAGAAACGAGCTGTC 2534
845 aThrTyrAsnLeuThrLeuGlnTyrThrValAspLeuValAspSerAsn 862
2535 GTCCTTAGAGAGTACTGTACATCACTAGTGGCAGATGTCATCAAGAAC 2584
862 roAspCysThrThrThrLeuArgIleSerGlyAspSerTyrPheThrPhe 878
2585 CTGACTGCAGACAGCTCTCTTAACCAATACCTCGTGGAAAAACATACA 2634
879 GlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaGlyAsnHisP 895
2635 GGAAGCAATCTCAAGACAGCTGATGGAAGACAGAGGATCTTTTA 2684
895 eCysPheAsnSerAsnPheGlnAlaPheSerGlnPheSerPheGlnLeu 912
2685 TGCTCTCTCCAAATCTGAGGTCAACACTATCTATGGAATATC 2734
912 rGlyLysSerArgAsnTyrAsnValAspLeuGlyAlaLysTyrGlnPhe 928
2735 GTGATCTTCACGACTCAATGACAGATCTTGAGGTAGTTCAGCTTC 2784

```

seq_name: /SIDS6/gcgdata/geneseq/geneseq/NA1999.DAT.X06823

seq_documentation_block:

ID X06823 standard; DNA; 2793 BP.

AC X06823;

DT 26-Apr-1999 (first entry)

DE Chlamydia pneumoniae surface exposed protein Omp11 DNA.

KM Omp11; Outer membrane protein 11; surface exposed protein; antigen;

XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.

OS Chlamydia pneumoniae.

PN W09858953-A2.

PD 30-DEC-1998.

PF 19-JUN-1998; 98WO-DK00266.

PR 23-JUN-1997; 97DK-0000744.

PA (BIRK/) BIRKELUND S.
(CHR1/) CHRISTIANSEN G.

PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
Mygind P;

DR WPI, 1999-105610/09.

P-PSDB; W88424.

Species-specific test for identifying mammals infected with
Chlamydia pneumoniae - comprises detecting antibodies specific for
outer membrane proteins of C. pneumoniae or nucleic acids encoding
these proteins

Claim 6; Page 62-63; 115pp; English.

This DNA sequence codes for the novel 97.6 kDa surface exposed
protein Omp11 (see W88424) of the human respiratory pathogen
Chlamydia pneumoniae. By generating antibodies against C.
pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
was obtained which reacted with outer membrane proteins. The
antibody was used to identify the genes (see X06816-27) encoding
C. pneumoniae proteins (see W88417-28) in an expression library of
C. pneumoniae DNA. The genes are situated in 2 gene clusters:
Omp12, 11, 10, 5, 4, 13 and 14 in one cluster and Omp6, 7, 8, 9 and 15 in
the other, and encode polypeptides of about 89, 6-100.3 kDa and
about 56.1 kDa. The invention provides a new species specific test
for identifying mammals (including humans) infected with Chlamydia
pneumoniae. The test comprises detecting antibodies specific for
Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
membrane proteins, especially by PCR. The proteins are also used
in the diagnosis of C. pneumoniae infection in mammals. The
nucleic acids and proteins can also be used in the immunization of
mammals, the nucleic acids being particularly useful as DNA
vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.

Sequence 2793 BP; 803 A; 681 C; 567 G; 742 T; 0 other;

alignment_scores:

Quality: 1927.00 Length: 953
Ratio: 2.706 Gaps: 17
Percent Similarity: 74.711 Percent Identity: 43.442

alignment_block:

US-09-428-122-2 x X06823

Align seg 1/1 to: X06823 from: 1 to: 2793

```

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePyr 17
  |||||
  1 ATGAAATATACCTTGCACAAACCTCTGATCTTCGACTCTTGCTACTCC 50
17 GlnSerMet...IleAlaThr.....GluThrValLeuAspSer 30
  |||||
  51 CATTATATTAGCATGCACTTACGAGCAGCATGCTTCTTATCCCTTA 100
30 erAlaSerPheAspGlyAsnLysAsnGlyAsnPhSerValArgGlySer 46
  ::|||
  101 CAGATAGCTTGTGATGAGCGCGGCTCTACATTACTCCAAATCTCA 150
47 GlnGluAspAlaGlyThrThrThrThrLeuPheGlyAsnValThrLeu 63
  ::|||
  151 GCAGATGCCAATGGACGAACTGCTTATCAGGAAATGCTATATA.. 198
63 uAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsnAsp 80
  |||||
  199 .AACGATGCTGGGAAAGGCACACATTACAGCTGCTCTTACAGAA 247
80 hrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPheGln 96
  |||||
  248 CTACGGGTGATCTGACATTTACTGAAAGGATCACTTTCATTCAAC 297
97 ThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSerValVal 113
  |||||
  298 ACGTAGATGGGGTTCGAATGACAGAGCTGGGCA...AGCACACTGC 344
113 LAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleAla 130
  |||||
  345 TGATTAAGCCCTTACATTTCACAGATTCTTCAACCTTCTTCAATGAG 394
130 erProGlySerSerIleThrThrGlyLysGlyAlaValSerCysSerThr 146
  ::|||
  395 CTCCTGACACTACAGTGTGCTCAGAAAAGACTTTAG...TTGCA 441
147 GlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSerLysAsnPh 163
  |||||
  442 GGAGCCTTAATCTTACCGATATGAACTGCTCTTATGACCAAAACGT 491
163 eSerThrAsp.....AsnGlyGlyAlaIleThrAlaLysThrLeu 177
  |||||
  492 CTCATGATGAGTATATACATGCGGAGCGATCCACCAAAACCTTT 541
177 erLeuThrGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSer 193
  |||||
  542 CTATTCTGGGAATACCTCTTCTAATACCTTCACTAGTATATACGCAAA 591
194 LysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrLys 210
  |||||
  592 AAATTAGGTGAGGAGCATCTAGCTCGCGCTGCAATGTTTCAGAA 641
210 nGlnGlyGluValSerPheSerAspAsnThrSerAspSerGlyAla 227
  |||||
  642 CACGCGCAGTACTCTTATGATAATAAGAGAACTGGGGGGGG 691
227 IaIlePheThrGluAlaSerValThrIleSerAsnAsnAlaLysValSer 243
  |||||
  692 CTCGCGGCTTGAAGCAGCTCTCGATTACTCAAAATAGCTCTTTC 741
244 PheIleAspAsnLysValThrGlyAlaSerSerThrThrGlyAspMe 260
  |||||
  742 TTCTCTGAAACACTGCACACAGATGCTCAGGC..... 774
260 tSerGlyGlyAlaIleCysAlaThrLysThrSerThrAspThrLysVal 277
  ::|||
  775 .AAGGCGGGCCATTATTGTAAGAAAAACAGAGAGACTCTACTCTTA 823
277 hrLeuThrGlyAsnGlnMetLeuLeuPheSerAsnAsnThrSerThr 293
  |||||

```

824 CTATCTGGAAATAAAGTCTGACCTTCGCCGAGAACTCTTCACTACT 873
294 AlaglyglyalalIeTyrrVallysLysLeuGluleuAlaSerGlyGlyLa 310
874 CAAGCGGAGCAATCTGTGCTGATGTCTAGATCTTCCGCTGCGGCC 923
310 uThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaProlyssGly 327
924 TACCTATTTCATAAATAATAGATGCGGAGCAACGCTGCAGGAGGCG 973
327 LysAlaIleAlaIleGluAspSerGlyGlyLeuSerLeuSerAlaAspSer 343
974 GCGCATATGCAATTCGCACTCTGATCTTAAGTCTCTCGCAAAATGAA 1023
344 GlyAspIleValPheLeuGlyAsnThrValThrSerThrThr...ProG 359
1024 GGAGCAATCAGCTTCCTTGGCAACACTTAACCTCAACCTCCGCCCAAC 1073
359 YThrAsnArgSerSerIleAspLeuGlyThrSerAlaLysMetThrAla 376
1074 ATCAGACAGGATGCTATCTACTGCGATGCTGCAGCAAAATTCAGAACT 1123
376 euArgSerAlaIleGlyArgAlaIleTyrrPheTyrrAspProle... 390
1124 TAAGGCGAGCCCAAGCCCAATCTATCTATTCATGATCCGATGCACT 1173
391 ...ThrThrGlySerSerThrThrValThrAspValLeuLysValAsnG 406
1174 AACACCCAGAGGCTCA.....GACGTCGACCATCAACCA 1211
406 uThrProAlaAspSerAlaLeuGlnTyrrThrGlyAsnIleIlePheThr 423
1212 ACCGATATGCAACATCGCTTAGATATTCAGAAAGATGTATTTCTG 1261
423 LysGlyLysLeuSerGlyuThrGluAlaAlaAspSerLysAsnLeuThrSer 439
1262 GGGAAAGCTCTTCAGATGACAGCAAGCTGCTGATTAACCTTCACACT 1311
440 LysLeuLeuGlnProValThrLeuSerGlyGlyThrLeuSerLeuLysH 456
1312 ATATTAAAGCAACATTCGCTCTGACCTCTGAACTTACGACCTCAAAAG 1361
456 sGlyValThrLeuGlnThrGlnAlaPheThrGlnAlaAspSerArgL 473
1362 AAATGTCAGATGATGTCATGTTTCACACAGCTGAGCTGCTACAC 1411
473 euGluMetAspValGlyThrThrLeuGluProAlaAspThrSerThrIle 489
1412 TCTCATGCAACCAAGCAAAAGCTCAAA...GCAGATCTGAAGCTATGC 1458
490 Asn.....AsnLeuValIleAsnIleSerSerIleAspGlyAlaLys 504
1459 AGCTTACCAACATTCGCTGATCTTCTGCTTAGAGGAATAAAG 1508
504 sAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerGlyT 521
1509 TGTGTCAATTAACAGCAGGAGCAACAAATATACTACTAACCTGC 1558
521 hrIleThrLeuAspProThrGlyThrPheTyrrGluAsnHisSerLeu 537
1559 CTCTTGTTTTCAGATATGAGCGCAATTTTATGAAGGCAATACGATA 1608
538 ArgAsnProGlnSerTyrrAspIleLeuGluLeuLysAlaSer..... 551
1609 AACCAAGCTTCACGAGCCCTTGTTGATATCACTGCTACTGCTGC 1658
552GlyThrValThrSerThrAlaValThrProA 562
1659 TAGCGATATTATATCGATGCGCTTCACACTTCACAGTACAAACTGCAG 1708
562 sProIleMetGlyGlyLysPheHisTyrrGlyGlnGlyThrIleGly 578
1709 AACCT.....CATACGGGATTCAGGGACATTGGGA 1740

579 ProIleValThrGlyThrGlyAlaSerThrThrAlaThrPheAsnThr 595
1741 GCCACTTGGGAGACATCACTCAAAATCAGAGCAATGACTTGCTGGT 1790
595 rLysThrGlyTyrrIleProAsnProGluArgIleGlySerLeuValPro 612
1791 AACTACGGGCTACACCACTTAATCTGAGCGTAGAGCTTCGTAAGTCCG 1840
612 snSerLeuThrPasnAlaPheIleAspIleSerSerLeuHisTyrrLeuMet 628
1841 ATCATATTATGGCATCTCTTACTAGCATTCGACACTACAGCAAGATCATG 1890
629 GluThrAlaAsnGluGlyLeuGlnGlyAspArgAlaPheThrProysAla 645
1891 ACATCTCAAGCGAATAGATATCTATCAGCAACGAGGACTCTGGGCTACAG 1940
645 YLeuSerAsnPhePheHisLysAspSerThrLysThrArgGlyPheA 662
1941 AACTGCCAATTTCTTCATAGGATAATATCAGCAACTACCAAGCATTC 1990
662 rGHisLeuSerGlyGlyTyrrValIleGlyLysAsnLeuHisThrCysSer 678
1991 GACATAAAGCTACGCTATATGCTTGAGCAAGTCTGAAGATTTTCT 2040
679 AspLysIleLeuSerAlaAlaPheCysGlnLeuPheGlyArgAspArg 695
2041 GAAATATCTTCACTAGTATGCTTCTGCACTCTCGTGGTAAAGATAAGA 2090
695 PThrPheValAlaLysAsnGlnGlyThrValTyrrGlyGlyThrLeuTyrr 712
2091 CCGTTTATATGTTGAATAATACCTCTCATACTATTAGCTTCCTATACC 2140
712 yGlnHisAsnGluThrTyrrIleSerLeuProCysLysLeuArgProCys 728
2141 TGCACATTCAGCATTCCTAGGAGACTTCC.....ATGCCCTCA 2181
729SerLeuSerTyrrValProThrGluIleProValLeuPheSerG 743
2182 TTGGAAGTATCACGCAATGCTGAAGATATTCCTCTCTTTGAAATGC 2231
743 yAsnLeuSerTyrrThrHisThrAspAsnAspLeuLysThrLysTyrrH 760
2232 CCAAGCTACGCTACAGCTCACTAAATGATATGATAGTACCTGCTATCT 2281
760 hrTyrrProThrValLysGlySerThrGlyAsnAspSerPheAlaLeuGlu 776
2282 CTATCTGAGCTCAAGTTCCTTGACCAATATCTGGGCTCTAGAG 2331
777 PheGlyGlyArgAlaProIleCysLeu...AspGluSerAlaLeuPheG 792
2332 CTGCGAGAGATCTGCGCTATATCTCCCTTAAGAACACACCTGCTTCCA 2381
792 uGlnTyrrMetProPheMetLysLeuGlnPheValTyrrAlaHisGlnGlu 809
2382 GGGATATTTCCCTTCTTAAGTTCAGCGAGCTCTACAGCGCCCAACAA 2431
809 LysPheLysGlnGlnGlyThrGluAlaArgLysPheGlySerSerArgLeu 825
2432 ACTTAAAGAGAGTGGCGCTGAAGCCGCTGTTGATGATGAGAGCTTA 2481
826 ValAsnLeuAlaLeuProIleGlyIleArgPheAspLysLysLeuAsp 842
2482 GTAGACTGCTATCCCTGCGCATTCGTTGAAATAATCCGAGAG 2531
842 sGlnAspAlaThrTyrrAsnLeuThrLeuGlyTyrrThrValAspLeuVal 859
2532 TGAATAAATAATTTGAGATTTCTTAGCAACATTTGATGATGATATC 2581
859 rGSerAsnProAspCysThrThrThrLeuArgIleSerGlyAspSerThr 875
2582 GTAAATAATCCCGTTCGCTACTTCTTAATGCTAGTGAGGCTCTTGG 2631

```

876 LysThrPheGlyThrAsnLeuAlaArgGlnAlaLeuValLeuArgAlaG1 892
      ::::: |||||
2632 ACTTGCGTATGTAACCTCGCAGACAGCTCTTAGCAGAGCTCG 2681
      ::::: |||||
892 YasnHisPheCysPheAsnSerAsnPhGluAlaPheSerGlnPheSerp 909
      ::::: |||||
2682 AAGCATCTGACTCTCTCCCATGTAGAACTCTCTGGGAGAGCTGCTT 2731
      ::::: |||||
909 heGluLeuArgLysSerSerArgAsnTyrAsnValAspLeuGlyAlaLys 925
      ::::: |||||
2732 ATGACCTCGTGGCTCAGCAGCATCTCAAAATGATGTTGGGCTAGA 2781
      ::::: |||||
926 TyrGlnPhe 928
      ::::: |||||
2782 TACTCATTC 2790
      ::::: |||||

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.X06816

seq_documentation_block:
ID X06816 standard; DNA: 3200 BP.
XX
XX X06816;
XX
AC
XX
DT 26-APR-1999 (first entry)
DE Chlamydia pneumoniae surface exposed protein Omp4 DNA.
XX
XX Omp4; outer membrane protein 4; surface exposed protein; antigen;
KM infection; diagnosis; vaccine; atherosclerosis; aschma; ss.
XX
OS Chlamydia pneumoniae.
XX
FH key Location/Qualifiers
FT CDS 205..2991
FT FT /*tag= a
XX
XX MO9858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
XX
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Mygind P;
XX WPI: 1999-105610/09.
XX P-PSDB: W88417.
XX
XX Species-specific test for identifying mammals infected with
XX Chlamydia pneumoniae - comprises detecting antibodies specific for
XX outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins
XX
XX Claim 6; Page 35-40; 115pp; English.
XX
XX This DNA sequence codes for the novel 98.9 kDa surface exposed
XX protein Omp4 (see W88417) of the human respiratory pathogen
XX Chlamydia pneumoniae. By generating antibodies against C.
XX pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
XX was obtained which reacted with outer membrane proteins. The
XX antibody was used to identify the genes (see X06816-27) encoding
XX Omp4-Omp15 proteins (see W88417-28) in an expression library of
XX C. pneumoniae DNA. The genes are situated in 2 gene clusters:
XX Omp12, 11, 10, 5, 4, 13 and 14 in one cluster and Omp6, 7, 8, 9 and 15 in
XX the other, and encode polypeptides of about 89.6-100.3 kDa and
XX about 56.1 kDa. The invention provides a new species specific test
XX for identifying mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for

```

```

CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
XX
XX Sequence 3200 BP; 946 A; 708 C; 623 G; 923 T; 0 other;

alignment_scores:
Quality: 1855.00 Length: 949
Ratio: 2.650 Gaps: 20
Percent Similarity: 73.762 Percent Identity: 42.677

alignment_block:
US-09-428-122-2 x X06816 ..

Align seg 1/1 to: X06816 from: 1 to: 3200

1 MetLysSerSerPheProLysPheValPheSerThrPheAlaIlePhePr 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 ATGAGACTTCGATTCCTCGGTTTAGTTCTCCTCGCTTAGCTTTC 254
17 O.....LeuSerMetIleAlaThrGluThrValLeuAspSerSerAlas 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
255 ATGTCACTTACAGTCACTAGTACGAGAGACTTTTATACCTGATGARA 304
32 erPheAspGlyAsn...LysAsnGlyAsnPheserValArgLysGln 47
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
305 GCTTATATGGAATATCGATTCAGAGACTTACTCCAAACTTCA... 351
48 GluAspAlaGlyThrThrTyrLeuPheLysGlyAsnValThrLeuGluAs 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
352 .....GCCACACATATTCCTTAACGAGAGATGCTCTTTACGA 392
64 nileProGlyThrGlyThrAlaIleThrLysSerCysPheAsnAsnThrL 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
393 G...CTGGAAAAGGCACCTCCCTTATGACAGTGTGTTTAAGCAACCA 439
81 yscLysAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPheGlnThr 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
440 CGGACATCTTACCTCTGCGGAGCGATGACTTAACGTTGGCTT 489
98 ValAspAlaGlyThrValAlaGlyAlaValAsnSerSerValValAs 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
490 ATAGATGCTGGCAGCTCAGTCAGTGTCTGTCA...TCTACACAGCAAA 536
114 pLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleAlaSerp 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
537 TAGAATCTTACCTTCTCAGGTTTCTTACTGAGTTTGTATTCCTTC 586
131 rGlySerSerIleThrThrGlyLysGlyAlaValSerCysSerThrGly 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
587 CTAGCACAACGGTTACTACAGGTGAGGAGACGCTTCC...TCAGCAGA 633
148 SerLeuSerLeuThrLysAsnValSerLeuPheSerLysAsnPhese 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
634 GCGTAATATTAGAAATAATTCGTAACTGTAGTGTGGGAAATTTTC 683
164 rThrAspAsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyT 181
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
684 TACTCAGATGTGTGAGTATCAAGAGAGCTCTTCTTTAACTGGCA 733
181 hrThMeSerAlaLeuPheSerGluAsnThrSerSerLysLysGlyGly 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
734 CTTCGAGAGATGCTCTTTTAGTAACAACCTCATCATCAACAAAGGAGA 783
198 AlaIleGlnThrSerAspAlaLeuThrThrLeuThrLysGlnGlyGly 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
784 GCAATTGCTACTACGAGCGGCTCGCATAGCAATAAATACACAGGTTATGT 833

```


511 ahrtsrlysaenleuthrleuSerGlyThrIlethrleuLeuaspProt 528
1698 TGGTTCCAAATAAAGATGATCGTATGGGAAAGTCACACCTGATTGATATG 1747
528 hrcglythrphetrlyguasnHtsserleuArgaspProGlseryasp 544
1748 AAGGACACTTATGAAAGTCTATGTTCCAGCATGACCAGCTCTTCT 1797
545 lIleuGluLeuLysAlaSerGlyThrValThrSer.....Th 557
1798 CATTAAATCAACGGTATGCTGATGATGTGATCACTAACGTTGACATCAG 1844
557 rAlaValIthrProaspProIleMeGlyGluLysPheHisTyrglyTyrg 574
1848 CAGCCTTATCCCTGTTCCGCTGAGAGATCCTAATTCAGAAATCGGATTCC 1897
574 InglyThrProglyProIleValItrpGlyThrGlyAlaSerThr.. 589
1898 AAGACAAATGGAAT...GTATATGTGACATGAGATACAGCTCAAAATACA 1944
590AlaThrPheasnTrpThrIlyThrGlyTyrlleProaspProG 604
1945 AAGAGCCACGCACTGGACCAAAACAGGATTGTTCACAGCCCGA 1994
604 uArgIleGlySerLeuValProaspSerLeuTrpAsnaIaPheIleAspI 621
1995 AAGAAATGTGGGTAGTATGCAATACCTATGGGAGATCTTTACTGACA 2044
621 lserSerleuHisTyrlleuMetGluThrAlaasnGlyLeuGlnGly 637
2045 TTTCGCTCTGTGACACGCTGTGAGAGATCCGGCAACTGGATAGGACAC 2094
638 AspArgAlaPheTrpCysAlaGlyLeuSerasnPhePheHisLysAspSe 654
2095 AAGCAAGTTCGTGGTTCTCCATGACAGACTTCCTGCATAGACATCG 2144
654 rThrIlyThrArgArgGlyPheArgHisLeuSerGlyGlyTyrValIleG 671
2145 AGATGAAATCCCAAGGCTCCGTCATACCTGTGAGAGCTACGATCGATCG 2194
671 lGlyIleuLeuHisIhrPrcysSerAspLysIleLeuSerAlaIaPheGys 687
2195 GGGGAAGTCTCACACTCTTAAGACAGACTATTTACCTTTGCGTTCG 2244
688 GlnLeuPheGlyArgAspArgAspIlyrPheValAlaLysasnGlnGlyTh 704
2245 CATCTGTTGCTAGAGACAAAGATTGTTTATCGCTCAACAACTACAG 2294
704 rValTyrglyGlyThrLeuTyTrpGlnHisasnGluThr..... 717
2295 AACCTACGGTGAAGACTTTATTTCTAACCACTCATACCTACAAACCC 2344
718TyrIleSerLeu...ProCysLysLeuArgProCysSerLeuSer 731
2345 AAACGATTTTGAGATTAGGAAGAAGCAAGATTTCGATGATCAGCTTAAGA 2394
732 TyrValProThrGluIleProValLeuPheSerGlyLysnLeuSerTyTh 748
2395 AATTCTCCAGGAAATTTCCCTCACCTCTGGATGTCACAAATTCCTGTAG 2444
748 rHisThrAspAsnAspLeuLysThrLysTyThrThrIlyrProThrVal 765
2445 CCATTGACCAACCGTATGGAACAGCACTATATCCATTCATGGCAGAAATCG 2494
765 uGlySerTrpLysAsnAspSerPheAlaLeuGluPheGlyValTyrgAla 781
2495 AAGGTCTGTGGAGCAACGAGTATAGCTGGGATGAGGCTACATCCT 2544
782 ProIleGlyLeu...AspLysSerAlaLeuPheGlnGlnIlyrMetProPh 797
2545 CCTTTGTGTTCTTCCAAACCCACACTCTCTTTTCAAGACCTTCAATTCACA 2594
797 eMetLysLeuGlnPheValTyrgAlaHisGlnGluIlyrPheLysGlnG 814

[illegible][illegible]

```

2416 GAAACCAATACACGACATATCCAGAAAGCTCAGGATCTGGGCAATG 2465
771 spSerPheAlaLeuGluPheGlyAlaPheProIleCysLeuAspGlu 787
2466 ATGTTTGTGCTGTGAGTTTGAGCGACTACTACTACTACCTAACAGT 2515
788 SerAlaLeuPheGluIntYrMetProPheMetLysLeuGlnPheValTy 804
2516 ACTTTTATTTATTTACTCTCTCCGTTTCTCAGCTGCAGTGCACCTA 2565
804 rAlaHisGlnGluGlyPheLysGlnGlnGlyThrGluAlaArgGluPheG 821
2566 TGCTCACCAGAAAGACTTCAAGAGACAGAGAGTGCAGTTGCTACTT 2615
821 LysSerArgLeuValAsnLeuAlaLeuProIleGlyIleArgPheAsp 837
2616 CTAGCGGAGATCTTTTCAATTTACAGCTTCTATTTGCGTGAAGTTGAG 2665
838 LysGluSerAspCysGlnAspAlaThrTyrAsnLeuThrLeuGlyTyrTh 854
2666 AGATTTTACAGACTGTAAAGGGGATCTTATGACTTACCTTGCTTATGT 2715
854 rValAspLeuValArgSerAsnProAspCysThrThrLeuArgIles 871
2716 TCCTGATGTGATTCGCAAGATCCCAAGACAGCGCAACATTTG...GCTA 2762
871 erGlyAspSerTyrPlyThrPheGlyThrAsnLeuAlaArgGlnAlaLeu 887
2763 GTGAGAGTACGTGAGACACCCAGGAACATCTCTCCACACAGGATTA 2812
888 ValLeuArgAlaGlyAsnHisPheCysPheAsnSerAsnPheGluAlaPh 904
2813 CACTGCGGTTAGGAAACCACTGTCTCATTAATCTCGAATTAGGCGTT 2862
904 eSerGlnPheSerPheGluLeuArgGlySerSerArgAsnTyrAsnVala 921
2863 CACTCACGGAGATTGAATTGCGGGATCTCTCGTAATTATACATCA 2912
921 splLeuGlyAlaLysTyrGlnPhe 928
2913 ATCTCGGGGTAAATACCGATTT 2935

seq_name: /SIDS6/gcdata/geneseq/geneseq/NA1999.DAT:X06827
seq_documentation_block:
ID X06827 standard; DNA; 2838 BP.
AC X06827,
XX
XX
XX 26-APR-1999 (first entry)
XX
XX Chlamydia pneumoniae surface exposed protein Omp15 DNA.
XX
XX Omp15: outer membrane protein 15: surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX Chlamydia pneumoniae.
XX
XX MO9858953-A2.
XX
XX 30-DEC-1998.
XX
XX 19-JUN-1998; 98WO-DK00266.
XX
XX 23-JUN-1997; 97DK-0000744.
XX
XX (BIRK/) BIRKJUND S.
XX (CHR/) CHRISTIANSEN G.
XX
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Myind P;
XX

```

WPI: 1999-105610/09.
P-PSDB: W88428.
Species-specific test for identifying mammals infected with Chlamydia pneumoniae - comprises detecting antibodies specific for outer membrane proteins of C. pneumoniae or nucleic acids encoding these proteins
Claim 6; Page 70-71; 115pp; English.

This DNA sequence codes for the novel surface exposed protein Omp15 (see W88428) of the human respiratory pathogen Chlamydia pneumoniae. By generating antibodies against C. pneumoniae outer membrane complex, a polyclonal antibody (PAB 150) was obtained which reacted with outer membrane proteins. The antibody was used to identify the genes (see X06816-27) encoding Omp4-Omp15 proteins (see W88417-28) in an expression library of Chlamydia pneumoniae DNA. The genes are situated in 2 gene clusters: Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in the other, and encode polypeptides of about 89.6-100.3 kDa and about 56.1 kDa. The invention provides a new species specific test for identifying mammals (including humans) infected with Chlamydia pneumoniae. The test comprises detecting antibodies specific for Omp4-Omp15 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of C. pneumoniae infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with C. pneumoniae.

Sequence 2838 BP; 837 A; 591 C; 648 G; 762 T; 0 other;

alignment_scores:
Quality: 1811.00 Length: 945
Ratio: 2.621 Gaps: 24
Percent Similarity: 73.122 Percent Identity: 42.963

alignment_block:
US-09-428-122-2 x X06827 ..

Align seg 1/1 to: X06827 from: 1 to: 2838

```

22 AlathrGluThrValLeuAspSerAlaSerPheAspGlyAsnLysAs 38
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 GCCGCAACCACTCCACATAATCTGAAGATGGTTTATTTGGGAGCGCAA 116
38 nGlyAsn...PheSerValArgGluSerGlnGluAspAlaGlyThr 54
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 TACAATATCTTTTCTCCGAATCTACACAGGATGTCGAGAACTACCT 166
54 yrLeuPheLysGlyAsnValThrLeuGluAsnIleProGlyThrGlyThr 70
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 ACTCTCTCACAGCAGAGGTTCTGTATAGAT...CCGGGAAAGGTGT 213
71 AlaIleThrLysSerCysPheAsnAsnThrLysGlySplLeuThrPhe 87
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 TCAATTACAGGAAGCTGCTTGTAGAAAGTGCAGCATCTTCAATTTT 263
87 rGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspAlaGly...Thr 103
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
264 AGTAAATGAAATACCTTAAGTTCTGTCGGTAGAGCGAGGCTATATA 313
103 alAlaGlyAlaAlaValAsnSerSerValValAspLysSerThrPhe 119
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 TCGCGTGTCTCATGTACAAAGAGT.....AAGAATTAAGCTTC 354
120 IleGlyPheSerLeuSerPheIleAlaSerProGlySerSerIleThr 136
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 ACAGATTCTCTCTGTGATCACAGAAATCTCAAAATCCGCTGTAG 404

```

```

136 rThrglyGlyAlaValSerCysSerThrGlySerLeuSerLeuThrL 153
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
405 TACAGGAAAGTAGCCTA...GTGAGTTCAGGAGCAGTCCACTGCAAG 451
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
153 ysaValSerLeuLeuPheSerLysasnPheSerThrAspAsnGlyLy 169
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
452 ATATAACACACTGTTCTTACAGCAATGCCCTGTCGAAAGATGGGCG 501
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
170 AlaIleThrAlaLysThrLeuSerLeuThrGlyThrMetSerAlaLe 186
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
502 GTGATTAAAGAAACTCCTGCTGATTACAGGAAATCAAAATAGTCGCA 551
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
186 upPheSerGluasnThrSerSerLysGlyAlaIleGlnThrSera 203
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
552 TTTGGCAAAATACATCTTCGAAAAAGAGGGCGCATCTCCAGCACTC 601
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
203 spAlaLeuThrIleThrGlyAsnGlnGlyValSerPheSerAspAsn 219
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
602 AAGGACTCACCATAGAGAAATTAAGGAGCGTAAAGTTCAATGAAGAAC 651
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
220 ThrSerSerSergLyAlaAlaIlePheThrGluAlaSerValThrI 236
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
652 AAACAGCATGACCTCAGAGCGCGCTAGATTAGAGCCCGCTCACTAT 701
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
236 eSerAsnAsnAlaLysValSerPheIleAspAsnLysValThrGlyAla 253
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
702 CACGCGCAACCATAGATTGATATTTTCACAAATTAAGACTTCTGGGATG 751
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
253 eSer.....SerThrGlyAspMetSer..... 261
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
752 CTGCAATGCGGAGCCATTAATTGCTCAGCGGACCTAACATTTACTGAT 801
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
262 .....GlyGlyAla 264
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
802 AACACTCTTTGTTACTTCAAGAAATAGCACATGCAAGATGGTGGAGC 851
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
264 aIleCysAlaTyrlLysThrSerThrAspThrLysValThrLeuThrGly 281
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
852 TTTGTGT.....AGCAGAGAAC...ATAAGCATTTACCGGTA 886
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
281 snGlnMetLeuLeuPheSerAsnAsnThrSerThrThrAlaGlyAla 297
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
887 GTGATTTCATCAATGTATAGAAATACCTTCAGACAAAAAGAGGAGCG 936
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
298 IleTyrlValLysLeuGluLeuAlaSerGly...GlyLeuThrLys 313
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
937 ATTTCCTCAGCTTCTCTCAAGATTTTGGAGGCGAGGAGCGCTCTCT 986
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
313 eSerArgAsnSerValAsnGlyGlyThrAlaProLysGlyAlaIleA 330
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
987 TTTAAATAAGTAGTACTCATGCCACC...CCTCTAGAGGTGCCATTT 1033
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
330 lalIleGluAspSergLyLeuLeuSerLeuSerAlaAspSergLyAsp 346
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1034 TTATCAACACAGAGGAGATCTTCGAGCTTCTCACTCAAGGAGGGATATC 1083
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
347 ValPheLeuGlyAsnThrValThrSerThrProGly.....ThrAs 361
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1084 GTATTCAGGAGGATCAGGTCACTACACAGCTCAATGCTTACACATA 1133
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
361 nArgSerSerLeaSpLeuGlyThrSerAlaLysMetThrAlaLeuArg 378
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1134 GAGAAATGTAAATTCACCTCGAGACACCGGCAAGTGAAGGAGTCTGCT 1183
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
378 eAlaAlaGlyArgAlaIleTyrlPheTyrlAspProIleThrThrGlySer 394
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1184 CAAGTCAAGTACGCTATCTATTTCTATGATCCCATTTACCAAC...AAC 1230
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
395 SerThrThrValThrAspValLeuLysValAsnGlyThrProAlaAsp 411
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1231 GATACGGGAGCAAGCAATACTTACATCAATAGAGTCACTGCAAAATCA 1280
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
411 rAlaLeuGlnTyrlThrLysAlaIlePheThrGlyLysLeuSerg 428

```

```

1281 : ||| : |||:|||||:|||||:|||||:|||||:|||||:|||||
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
428 luthrGluAlaAlaAspSerLysAsnLeuThrSerLysLeuLeuGlnPro 444
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1325 CAGAGAGAGCTATA...GCTGAATACTTCTCGAGGATCAACCAAGCCT 1371
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
445 ValThrLeuSerGlyGlyThrLeuSerLeuLysGlyValThrLeuG 461
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1372 GTCACCTTAGTAGAGGAGACCTTAGAACTTAAACAGAGAGTGCCTTGAT 1421
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
461 nThrAlaAlaPheThrGlnGluAlaAspSerArgLeuGluMetAspValG 478
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1422 CACACAGAGATTCTCGAGAGCCAGAGATCCAGCTTCTTTGGATTGG 1471
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
478 lYThrThrLeuGlnProAla...AspThrSerThrIleAsnAsnLeuVal 493
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1472 GGACCTCATACAAAGCTTTACAGAGATATGCTCATCACAAAATTCATCT 1521
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
494 lLeaSnLeSerSerIleAspGlyAlaLysLysAlaLysIleGlnThrLy 510
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1522 ATAAATGCCGATACCAATTAGGAAAGAAATCAATCAATATTTACTCTTC 1571
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
510 sAlaThrSerLysAsnLeuThrLeuSergLyThrIleThrLeuLeuAsp 527
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1572 AGCAGGCAATTAAGACATTAACCTTAACAGAGACCTTAAGCATTGTAATG 1621
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
527 roThrGlyThrPheTyrlAsnHisSerLeuArgAsnProGlnSerTyrl 543
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1622 CAGATGGAGCTTTGTATGAAACCATACCTTGCAAACCTCTCAAGATTAA 1671
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
544 AspIleLeuGlnLeuLys.....AlaSerGlyThrValThrSerThrAl 558
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1672 AGCTTGTAAAGTTATCTCCAGAGCGGAGGAGACTATA..... 1710
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
558 aValThrProAsp.....ProIleMetGlyLysAsp 569
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1711 .ATTACTCAGAGATGCTTCTCAGAACTTCTTGAAGTAGCTCTTACAGC 1759
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
569 henTyrlGlyTyrlGlnGlyThrTyrlGly...ProIleValTyrlGlyThr 584
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1760 CACATTATGGCTATCAAGGACATTGGAATGTGCAAGTCAACCCAGGAAAG 1809
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
585 GlyAlaSer...ThrThrAlaThrPheAsnThrPThrLysThrGlyTyrl 600
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1810 GGAACSTCAACGAGCCAGCAAGATTTAGAAATGGTGGCAGACAGATACCT 1859
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
600 eProAsnProGlnArgIleGlySerLeuValProAsnSerLeuTyrlAsn 617
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1860 TCCGAATCCGAGACGCAAGGATTTTATGTTCCCAATAGCCTGTGGGGT 1909
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
617 lArlleAspIleSerSerLeuHisTyrlLeuMetGlnThrAlaAsnGln 633
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1910 CTTTGTGTATCAGCGCTGATCCAAAGAAATCATGTGTAATATGATACCAA 1959
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
634 GlyLeuGlnGlyAspArgAlaPheTyrlCysAlaGlyLeuSerAsnPhe 650
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1960 ATCTTATGTAGAGACAGCGGAGCTGGGAGCTGGAATGCTATGCTATCC 2009
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
650 eHisLysAspSerThrLysThrArgArgIlePheArgHisLeuSerGly 667
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2010 ACATAGAGATAAAT...AATGAGCAGCGCTATCCCAATAGCGGTGG 2056
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
667 lYThrValIleGlyGlyAsnLeuHisThrCysSerArgLysLysLeuSer 683
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2057 GTTATCTGTGGGAGTGGCAGCTCATGCTTTTCTGATGCTAGCAATAAT 2106
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
684 AlaAlaPheCysGlnLeuPheGlyArgAspArgAspTyrlPheValAla 700
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2107 GCGGCTTTTGGCAGCTCTCAAGATGAAGAAAGTACAGTACAGTACCA 2156
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
700 sAsnGlnGlyThrValTyrlGlyGlyThrLeuTyrlGlnHisAsnGln 717
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

2157 AATTCATGAGACAGTACTACTACGGGGCTGATATTTCTTGAGATACCTAG 2206

717 htrYr1IleSerLeuProCysIysLeuArgProCysSerLeuSerTyVal 733
::: ||| :::
2207 AGTTTAGAGT...CCACAGGCAATCTTACTGATGATCTCTCCAGAACT 2253

734 ProthrGluIleProValLeupheSerCylAsnLeuSerTythrH1sth 750
::: ||| :::
2254 TGCCTAACCAAGTCGTACTATATATATGCAAGTGTCTTACAGCCATAG 2303

750 rAspAsnApLeuIysThrIysTyThrThrTyProthrValIysGly 767
:::|||||:::
2304 AATATATGATATGAAACCAATACAGACATATCCAGAGCTCAGAGAT 2353

767 eErTrpGlyAsnAspSerPheAlaLeuIupheGlyIupArgAlaIle 783
:::|||||:::
2354 CTGGGCAATGATGTTTGTGCTTACGTTTGAGGAGACTACATCTAC 2403

784 CysLeuAspGluSerAlaLeuPheGluIglTyMetProPheMetLys 800
:::|||||:::
2404 TACCTTAACAGACTTTTATTTGATTAATCTCTCCGTTCTCTCAGCT 2453

800 uGlnPheValTyraIahIsgIngluGlyPheLysGluInglTyH1gua 817
|||:::|||||:::
2454 GCAGTGCACATCTACTCTCCAGGAAGCTTCAAGAGACAGAGCTGAG 2503

817 lAerGluIupheGlySerSerArgLeuValAsnLeuAlaLeuProIleGly 833
:::|||||:::
2504 TTCTGCACATTTACTAGGGGAGATCTTTCAATTTTACAGATCTCTATTGGC 2553

834 lLeArgPheAspLysGluSerAspCysGlnAspAlaThrTyAsnLeuTh 850
:::|||||:::
2554 GTGAGATTTTGAGATTTTCCAGCTGTAAGAAGGATCTTAGAAGCTTAC 2603

850 rLeuGlyTyThrValAspLeuValArgSerAsnProAspCysThrThrT 867
|||:::|||||:::
2604 CCTTGCTATGTTCTCCATGTGATGTGCAAAATATCCCAAGACAGCGGCA 2653

867 hTrLeuArgIleSerGlyAspSerTrpLysThrPheGlyTyThrAsnLeuAla 883
|||:::|||||:::
2654 CATTG...GCTAGTGAGCTAGCTGAGGACACCACAGGAACAATCTCTCC 2700

884 ArgGlnAlaLeuValLeuArgIahGlyAsnH1SpheCysPheAsnSerAs 900
|||:::|||||:::
2701 AGCAACAGATTTACAATCTGCGTTTGGGAACCACTGCTCATATAAATCCTGG 2750

900 nPheGluAlaIleSerGlnPheSerPheGluIupArgGlySerSerArg 917
:::|||||:::
2751 AATTTGAGGTTCACGACGAGCATTAATGAATTTGGGGGAGATCTCTGTA 2800

917 snTyAsnValAspLeuGlyAlaLysTyArgInPhe 928
|||:::|||||:::
2801 ATTATATACATCAATCTCGGGGTAATATCCGATTT 2835

seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:X06819

seq_documentation_block:
ID X06819 standard; DNA: 2526 bp.
XX
XX X06819;
AC
XX
XX 26-APR-1999 (first entry)
DE Chlamydia pneumoniae surface exposed protein Omp7 DNA.
XX
XX Omp7; outer membrane protein 7; surface exposed protein; antigen
XX infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX Chlamydia pneumoniae.
XX
XX W09856953-A2.
XX
XX
XX 30-DEC-1998.
XX
XX

```

PF 19-JUN-1998; 98NC-DK00266.
XX
XX 23-JUN-1997; 97DK-0000744.
XX
XX (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
PI
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygind P;
DR WPI, 1999-105610/09.
XX P-PSDB: W88420.
XX
PT Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins
XX
XX
XX Claim 6; Page 49-50; 115pp; English.
XX
XX This DNA sequence codes for the novel 89.7 kDa surface exposed
XX protein Omp7 (see W88420) of the human respiratory pathogen
XX Chlamydia pneumoniae. By generating antibodies against C.
XX pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
XX was obtained which reacted with outer membrane proteins. The
XX antibody was used to identify the genes (see X06816-27) encoding
XX Omp4-Omp5 proteins (see W88417-28) in an expression library of
XX C. pneumoniae DNA. The genes are situated in 2 gene clusters:
XX Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
XX the other, and encode polypeptides of about 89.6-100.3 kDa and
XX about 56.1 kDa. The invention provides a new species specific test
XX for identifying mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for
XX Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
XX membrane proteins, especially by PCR. The proteins are also used
XX in the diagnosis of C. pneumoniae infection in mammals. The
XX nucleic acids and proteins can also be used in the immunization of
XX mammals, the nucleic acids being particularly useful as DNA
XX vaccines for effecting in vivo expression of antigens. The
XX vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.
XX
XX Sequence 2526 BP; 696 A; 495 C; 597 G; 738 T; 0 other;
XX
XX
XX Alignment_scores:
XX Quality: 1564.00 Length: 942
XX Ratio: 2.523 Gaps: 20
XX Percent Similarity: 65.817 Percent Identity: 39.172
XX
XX alignment_block:
XX US-09-428-122-2 x X06819 ..
XX
XX Align seq 1/1 to: X06819 from: 1 to: 2526
XX
XX 12 ThPha1a1lePheProLeuSerMetIleAlaThrGluThrValLeuAs 28
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 43 ACCCTTCTATGTCGAATTATTAGAGAGCTGTACCGAAGAGCTATC 92
XX
XX 28 pSerSerAlaSerPheaspGly...AsnYsaSncIYsaNpHeserYala 44
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 93 GCGTACCAATAGCTCGATGAGACTCATCACAACAGCTTTCTTAGTA 142
XX
XX 44 rgluSerGlnGluAspAlaGlyThrThyTyrLeuPheIYsaVal 60
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 143 AAACATCATCGGCTACAGATGCGACCAATATGTTTAAAGATTCGTA 192
XX
XX 61 ThIleGluAsnIlePheCgIYthrGlyThrAlaIleThrYsSerGlysh 77
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 193 GTTATAGAAATGTACCCAAACAGGGAACAGTCACTGCTACTGATGTT 242
XX
XX 77 easnAsn.....ThIYsGlyAspLeuThrPheThrGlyAsnGlyAsnS 92

```

```
243 TAAATAGACCGCTGACCTGGAGATCTAAATTTTTCAGAGGGGATTTT 292
92 erleuLeupheglnThrValAspAlaGlyThrValAlaGlyAlaAlaVal 108
CTTACATTTAGCATATCGATCGCAACGACGCTTCGGAGCTGCTAT 342
109 AsnSerSerValValAspLysSerThrThrPheIleGlyPheSerSerLe 125
GGAAGTACAGCAGCTAATAGACAGTCAGCTTATCAGCATTTTCGGACT 392
343 GGAAGTACAGCAGCTAATAGACAGTCAGCTTATCAGCATTTTCGGACT 392
125 userPheIleAlaSerProGlySerSerIleThrThrGlyLysGlyAlaVal 142
TTCTTTCTTAAATCCCAGCAGATACAGTACATGATGATGGAGACT 442
393 TTCTTTCTTAAATCCCAGCAGATACAGTACATGATGATGGAGACT 442
142 aLserCysSerThrGlySerLeuSerLeuThrLysAsnValSerLeuLeu 158
TCAATGTTAA...GGGAATTTAGCTTATGATGATGATGATGATGATG 489
443 TCAATGTTAA...GGGAATTTAGCTTATGATGATGATGATGATGATG 489
159 PheSerLysAsnPheSerThrAspAsnGlyValAlaIleThrAlaLysTh 175
ATTAGAGACATTTCTCAACGAGAGATGGGAGCAATTAATGT... 534
490 ATTAGAGACATTTCTCAACGAGAGATGGGAGCAATTAATGT... 534
175 rLeuSerLeuThrGlyThrMetSerAlaLeuPheSerGluAsnThrS 192
534 ..... 534
192 erSerLysLysGlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThr 208
534 ..... 534
209 GlyAsnGlnGlyValSerPheSerAspAsnThrSerSerAspSerG 225
534 ..... 534
534 ..... 534
225 yAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsnAlaLysV 242
535 .....GCAGGCTCCTTGAAGATCGCAACCAATAGTCC 568
242 aLserPheIleAspAsnLysValThrGlyAlaSerSerSerThrThrGly 258
569 TTTCTTTTAT... 579
259 AspMetSerGlyGlyAlaIleCysAlaTyrLysThrSerThrAspThrLy 275
579 ..... 579
275 sValThrLeuThrGlyAsnGlnMetLeuPheSerAsnAsnThrSerT 292
580 .....GGAATAGTCTT 592
292 hrThrAlaGlyValAlaIleTyrValLysLysLeuGluLeuAlaSerGly 308
593 CAACACGTGGCGGAGCATTCATACCAAAACCTCACACATATCTTCTGT 642
309 GlyLeuThrLeuPheSerArgAsnSerValAsnGlyLysThrAlaPro.. 324
643 GGGGAACCTCATTT...CAGGGGAATACAGCGCTTAC 677
325 .....LysGlyGlyAlaIleAlaIleGluAspSerGlyLysLeuS 338
678 GGCTGCTGTGAAGAGGTCTATCGGATTGACAGCTCTGGCCCTAT 727
338 erLeuSerAlaAspSerGlyAspIleValPheLeuGlyAsnThrValThr 354
728 CCAATTTCTGAGACAGTGGCAGCATTTATCTTTGAAGCAATACATAGCA 777
355 SerThrThrProGlyThr...AsnArgSerSerIleAspLeuGlyThrSe 370
778 GCTACA...GGAACTGCTCTCATATGCTATGATTTAGAACTAG 821
370 rAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleTyrPheT 387
|||||.....|
822 CGCTAGATACCTGCTTACGTCCTGCGCAGACATACATATCTTT 871
387 yAspProIleThrThrGlySerSerThrThrValThrAspValLeuLys 403
872 ATGATCGATTTACTGTATACAGATCGACATCTGTCTGATGCTCAAT 921
404 ValAsnGluThrProAlaAspSerAlaLeuGlnTyrThrGlyAsnIleI 420
922 ATTAATAGCCCTGATCTGAGATTAACAAAGAGTATACGGGAACCATAGT 971
420 ePheThrGlyLysLeuSerGlnThrGluAlaAlaAspSerLysAsnL 437
972 CTTTTCTGGAGACAGCTCAGCAGCAGACAGTAAAGATGAAAGAACACC 1021
437 eutThrLysLeuLeuGlnProValThrLeuSerGlyLysThrLeuSer 453
1022 GCATCTTAATTAATCTCAAAATGTTGCTTTTAAATAGGAGCATGATG 1071
454 LeuLysIleGlyValThrLeuSerGlnThrGluAlaPheThrGlnIleAla 470
1072 TTAAGAGTATGCTGTTTAAGTGGACGAGGATTTCTCAGATGCAAA 1121
470 pSerArgLeuGluMetAspValGlyThrThrLeuGluProAlaAspThrS 487
1122 CTCTAGGTGATGATGATTTAGGAGCGTGTG...GTTCGAACACCG 1168
487 erThrIle.....AsnAsnLeuValIleAsnIleSerSerIleAspGly 501
1169 AAAGTATCGAGTTAAGCAATTTGGAATTAATTAATTAATTAATTAATTA 1218
502 AlAlaLysAlaLysIleGluThrLysAlaThrSerLysAsnThrLeu 518
1219 GGGAAAAAGATTAAGTCACTGCTGCCACAGCTCAGAAAGATTTGCTAT 1268
518 uSerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGluAsnH 535
1269 AGATCGTCTGTTGCTGCAATTAAGTCAATTAAGTCAATTAAGTCAAAATG 1318
535 lAserLeuArgAsnProGlnSerTyrAsp...IleLeuGluLeuLysAla 550
1319 GCTTTTGAATGAGACACATCTCTATGATGGGATCTTCTGATTAAGTCT 1368
551 SerGlyThrValThrSerThrAlaValThrProAspProIleMetGly 567
1369 GGAAGACATCGTGAATTTCTGACATCTCGCAGT...ATAATCTGT 1415
567 uLysPheHisTyrGlyTyrGlnGlyThrTrpGlyProIleValTrpGlyT 584
1416 ACAATCTCCGTATGCTATCAGGGAAGTGG...ACAATCAATTTGGTCTA 1462
584 hrGlyAlaSerThrThrAlaThrPheAsnThrPheThrGlyTyrIle 600
1463 CTAT...GATTAAGAAAGCTACGCTTTCTTGGCAAGCAAAATTTTAAT 1509
601 ProAsnProGluArgIleGlySerLeuValProAsnSerLeuTrpAsnAl 617
1510 CCCACTGCTAGCAGGAGGCTCCGTAGTCTTAATCTTCTTGGGCTTC 1559
617 aPheIleAspIleSerSerLeuHisTyrLeuMetGluThrAlaAsnGlu 634
1560 TTTTATGATGTTGTCCTCCCTTCCAAATTTTATAGAGCTAGTACGAAG 1609
634 LysGlnGlnLysArgAlaPheThrCysAlaGlyLeuSerAsnPhePhe 650
1610 GGCTCTCTTCAAAAGAGATTTTGGTGGTACAGCATTTCCATGTTTGG 1659
651 HisLysAspSerThrLysThrArgArgGlyPheArgHisLeuSerGly 667
1660 CATAGAGCGGTGTGTAATAATCAAGAAATTCCTCATGTGAGTGAAG 1709
667 yThrValIleGlyLysLeuHisThrCysSerAspLysIleLeuSerA 684
1710 TGCTGTAGTAGTGTACGACGAGATGCCGGGTGTGATACCTTGTCTC 1759
```

```

684 laalaphcysglneuphegllyaraspargasptyrphevalalays 700
    ::::::::::::::::::::::::::::::::::::::::::::::
1760 tgggttttctcagctctttgcccgtgacaaagactcttattgaatacc 1809
    ::::::::::::::::::::::::::::::::::::::::::::::
701 aenglnglythvaltyrlyglythrlentyrtyrglnhlsasgluth 717
    ::::::::::::::::::::::::::::::::::::::::::::::
1810 aatttcgcaaaagacctacgacagctttacgttgacacagatgcttc 1859
    ::::::::::::::::::::::::::::::::::::::::::::::
717 tyrtlleserleuprocys.....lysleuathgproc 728
    ::::::::::::::::::::::::::::::::::::::::::::::
1860 cctatactctgtgagtagtctttagaagaggagactccgcagaga 1909
    ::::::::::::::::::::::::::::::::::::::::::::::
728 ysserleuserlyvalprothrgluileprovalleupheserglyasn 744
    ::::::::::::::::::::::::::::::::::::::::::::::
1910 tcctgttccttatgtttccaaagactcgcgcgtcttctatnoggcag 1959
    ::::::::::::::::::::::::::::::::::::::::::::::
745 leuserlytrhthsrthaspsaspleulysrthlystyr..... 758
    ::::::::::::::::::::::::::::::::::::::::::::::
1960 cttagctacggccatcagatcgcatagaagaccgagctctaccccc 2009
    ::::::::::::::::::::::::::::::::::::::::::::::
759 .....thrttyrprothrvallysglysertrpoglyasnspserp 773
    ::::::::::::::::::::::::::::::::::::::::::::::
2010 ccccccccgagcgcctcgcagcagatcattcttggagagatgctct 2059
    ::::::::::::::::::::::::::::::::::::::::::::::
773 healaleuqlupheglyglyaralargalproliacycleuaspluserala 789
    ::::::::::::::::::::::::::::::::::::::::::::::
2060 ggcttgagagacttggaactga.....gttgctgttgaataaacacagc 2103
    ::::::::::::::::::::::::::::::::::::::::::::::
790 .....leupheglnglntyrmetprohemetylsleuqlupheva 803
    ::::::::::::::::::::::::::::::::::::::::::::::
2104 ggcagagatattttccgagctacactcattgttaaaagccaaagctgt 2153
    ::::::::::::::::::::::::::::::::::::::::::::::
803 ltyralahisnglulgylphelysglngllythrglualearglup 820
    ::::::::::::::::::::::::::::::::::::::::::::::
2154 ttactccgccaagatagctttgttgaactagagagctatcagtcgatt 2203
    ::::::::::::::::::::::::::::::::::::::::::::::
820 heglysersearlgleuvalasneulaleuprollelylleargphe 836
    ::::::::::::::::::::::::::::::::::::::::::::::
2204 ttgctgttgcacattttatcacttcgattccttggaaatcaagttta 2253
    ::::::::::::::::::::::::::::::::::::::::::::::
837 asplysgluseraspcysglnaspaiathrtyrasneulhrleuqltyr 853
    ::::::::::::::::::::::::::::::::::::::::::::::
2254 gagaacagc.....tttgacagacatattatcattttacgagatga 2297
    ::::::::::::::::::::::::::::::::::::::::::::::
853 rthrvalaspleuvalargserasnproaspcysrthrthrlleuargt 870
    ::::::::::::::::::::::::::::::::::::::::::::::
2298 ttctccagatgtttgtctgtatgaacccaaatgtacgactaccccttt 2347
    ::::::::::::::::::::::::::::::::::::::::::::::
870 leserglyasertpysrthrlpheglythrasneulaaarglnala 886
    ::::::::::::::::::::::::::::::::::::::::::::::
2348 ccaaccagagagacttgagaaagccaaaggttcgaacttacgaaacagcgt 2397
    ::::::::::::::::::::::::::::::::::::::::::::::
887 leuvalleuarglaaglyasnhisphcysphesaserasnphleuul 903
    ::::::::::::::::::::::::::::::::::::::::::::::
2398 ggaattgttcagcgcctcaggttttcgattcttggcagcgcacagagct 2447
    ::::::::::::::::::::::::::::::::::::::::::::::
903 apbeserglupheserphegluleuarglyseraserargasnlyrasny 920
    ::::::::::::::::::::::::::::::::::::::::::::::
2448 tttcggggaactttggctttgaatggcgagatcttctctactataatg 2497
    ::::::::::::::::::::::::::::::::::::::::::::::
920 alaspplenglyalalyslyrglnphe 928
    ::::::::::::::::::::::::::::::::::::::::::::::
2498 tagatcgggtacgaaatcaaatattt 2523
    ::::::::::::::::::::::::::::::::::::::::::::::

```

```

seq_name: /SID56/gcgdata/geneseq/NA1999.DAT:X06818
seq_documentation_block:
ID X06818 standard; DNA; 3052 BP.
XX
AC X06818;
XX
DT 26-APR-1999 (first entry)

```

```

XX XX Chlamydia pneumoniae surface exposed protein Omp6 DNA.
DE XX
XX XX
XX Omp6: outer membrane protein 6: surface exposed protein; antigen;
KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX XX
XX Chlamydia pneumoniae.
XX XX
XX Key Location/Qualifiers
FH CDS 1..2764
FT /*tag= a
XX
XX MO9858953-A2.
XX
XX 30-DEC-1998.
XX
XX 19-JUN-1998; 98WO-DK00266.
XX
XX 23-JUN-1997; 97DK-0000744.
XX
XX (BIRK/) BIRKELUND S.
XX (CHRI/) CHRISTIANSEN G.
XX
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX Mylind P;
XX WPI: 1999-105610/09.
XX P-PSDB: W88419.
XX
XX Species-specific test for identifying mammals infected with
XX Chlamydia pneumoniae - comprises detecting antibodies specific for
XX outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins
XX
XX Claim 6; Page 45-46; 115pp; English.
XX
XX This DNA sequence codes for the novel 100.3 kDa surface exposed
XX protein Omp6 (see W88419) of the human respiratory pathogen
XX Chlamydia pneumoniae. By generating antibody against C.
XX pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
XX was obtained which reacted with outer membrane proteins. The
XX antibody was used to identify the genes (see X06818-27) encoding
XX Omp4-Omp15 proteins (see W88417-28) in an expression library of
XX C. pneumoniae DNA. The genes are situated in 2 gene clusters:
XX Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
XX the other, and encode polypeptides of about 89,6-100.3 kDa and
XX about 56.1 kDa. The invention provides a new species specific test
XX for identifying mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for
XX Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
XX membrane proteins, especially by PCR. The proteins are also used
XX in the diagnosis of C. pneumoniae infection in mammals. The
XX nucleic acids and proteins can also be used in the immunization of
XX mammals, the nucleic acids being particularly useful as DNA
XX vaccines for effecting in vivo expression of antigens. The
XX CC vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.
XX
XX Sequence 3052 BP; 875 A; 603 C; 653 G; 921 T; 0 other;

```

```

alignment_scores:
Quality: 1430.50 Length: 951
Ratio: 2.187 Gaps: 26
Percent Similarity: 68.770 Percent Identity: 36.698

```

alignment_block:

US-09-428-122-2 x X06818 ..

Align seg 1/1 to: X06818 from: 1 to: 3052

```

5 PhePro...LysPheValPheSerThrPheAlaIlePheProLeuSerMet 20
||||||| ||| ::::::::::::::| ||:::

```



```

22 TTCTCTACTTTTTTTTAACTGCTCTGAGTCTTGACACTTCTT 71
20 ttleAlaThrGluThrValLeuAspSerSerAlaSerPheAspGlyAsnL 37
   : ||||| ||| :||| ||||| |||||
72 GAGTGCACACTACGATTTCTTAACCCAGAAAGATAGTTTCATGCGAGATA 121
37 ysaAsnGlyAsnPheSerValArgGluSerGlnGluAspAlaGlyThr 53
   : ||||| :||| :||| :||| :||| :|||
122 GTCAGAAAT.....GCAGAACGTTCTATATATGTTCAAGCTGGGAGATGC 165
54 TyleuPheGlyGlyAsnValThrLeuGluAsnIleProGlyThrGlyTh 70
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
166 TATAGCTTCTGCTGATGCTCATATCTACGTC.....GATACTC 209
70 rAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspLeuThrPhe 87
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
210 TGCATTAAATAAAGCCTGCTTCAATGTCACCTCAGAAAGTGCAGCTTGC 259
87 hrcLysAsnGlyAsnSerLeuLeuPheGlnThrValAspAlaGlyThrVal 103
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
260 CAGGAATCATCATGGCTATATTATTTAATAATTTCTCAGGAACATACA 309
104 AlaGlyAlaAlaValAsnSerSerValAlaAspLysSerThrThr...Ph 119
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
310 AAGGAAGGCTGTACTTCTTGCCAAAGATCTTCAACCAACGACGACGTT 359
119 eileGlyPheSerSerLeuSerPheIleAlaSerProGlySerIleThr 136
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
360 TTCTGGTCTCCACGCTCTCTTTATTCAGAGCCCGAGATATTAAAG 409
136 hThrGlyLysGlyAlaValSerCys.....SerThrGlySerLeuSer 150
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
410 AACAGGA.....TGCTCTATTCAAAAATAATGCATTATG 444
151 LeuThrLysAsnValSerLeuLeuPheSerLysAsnPheSerThrAspAs 167
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
445 CTCTTAACATATATAGTAGCGTTTGAACAAACCAAGTAAGACTAA 494
167 nGlyAlaAlaIleThrAlaLysThrLeuSerLeuThrGlyThrMetLs 184
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
495 AGCGGCGATTAAGTGGGCGAATGTACTATAGTAGCACTACGAT 544
184 eAlaLeuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGln 200
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
545 CCCTCTCTTCTATCAGAAATGCAGCCACT...TTGGAGGTGCTATCCAT 591
201 ThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerPheSe 217
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
592 TCTTCAGATCCCTACAGATTGCAGTAATCAGGACAGATAAGATTGC 641
217 rAspAsnThrSerSerAsp...SerGlyAlaAlaIlePheThrGluAla 233
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
642 ACAAAATACGCCAGAAATGTTCTGGAGGGCTTTGTACTCCGATGCTG 691
233 eAlaThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysVal 249
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
692 ATATTGATATGATCAGAAATGCTTATGTTTTCAGAAATAGAC... 738
250 ThrGlyAlaSerSerThrThrGlyAspMetSerGlyAlaIleCys 266
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
739 .....GCATTGACTACTGCTATAGTAAGAGGAGGCTGCTG 776
266 sAlaIleThrLysThrSerThrAspThrLys.....ValThrLeuThr 279
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
777 TTGTCTTCCACTTCAGAGAGTAGTACTCCAGTCTTATTTGACTTCT 826
279 hrcLysAsnGlnMetLeuLeuPheSerAsnAsnThrSerThrThrAlaGly 295
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
827 CTGACATATAACAGATTAGTCTTGAAGAAGAACCATTCATTAATGGGGGC 876
296 GlyAlaIleThrValLysLysLeuGlnLeuAlaSerGlyLysLeuThrIle 312
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
877 GGAGCATTTATGCTAGGAACCTTAGCATCTTTCAGAGAGTCTTACTCT 926
312 upheSerArgAsnSerValAsnGlyThrAlaProLysGlyGlyAlaI 329
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
927 ATTATCATTAATATATCATATGCAAAATTCGCAAAATTAGGTGGAGCTTA 976
329 leAlaIleGluAspSerGlyGlyLeuSerLeuSerAlaAspSerGlyAsp 345
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
977 TTGCCATTATACTGAGGGGAGATCAGTTTATCAGACAGAAAGGAACA 1026
346 lIleValPheLeuGlyAsnThrValThrSerThrProGlyThrAsnAr 362
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1027 ATTACATTTCCAGGAAC.....CGACGAGCTTACCGTTTTCAT... 1068
362 gSerSerIleAspLeuGlyThrSerAlaLysMetThrAlaLeuArgSer 379
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1069 .....GCATTCACATCTTTACAAAATGCTAAATTCGAAATTACAGCGCA 1114
379 lAlaIleGlyArgAlaIleThrPheThrAspProIleThrThrGlySerSer 395
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1115 GAATGGAATGCTTATAGAATTTTATGATCTATTACT.....TCTGAA 1158
396 ThrThrValThrAspValLeuLysValAsnGluThrProAlaAspSerAl 412
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1159 GCAGATGGGTCTACCCCATTTGATATCAACGAGATCCTTAAATAATAA... 1206
412 aleuGlnThrThrGlyAsnIleIlePheThrGlyGlyLysLeuSerGluT 429
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1207 ...GAGTACACAGGAGCATTACTCTTTCTGGAGAAAG.....A 1243
429 hrcGluAlaAlaAspSerLysAsnLeuThrSerLysLeuLeuGlnProVal 445
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1244 GTCTACAAACAGATCCTAGGATTTAAATCTAACATTCCTCAGAACGTC 1293
446 ThrLeuSerGlyGlyThrLeuSerLeuLysHisGlyValThrLeuGlnTh 462
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1294 AACCTGTGCAAGAGATTAAGTATTAAAGAGGGCCGAGATGACAGT 1343
462 rGlnAlaPheThrGlnGlnAlaAspSerArgLeuGlnMetAspValGlyT 479
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1344 TTCAAATTCACGACACTCTCCAGGATTCGATTAAGTTAGATTGAGTA 1393
479 hThrLeu...GluProAlaAspThrSerThrIleAsnAsnLeuValIle 494
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1394 CCAAACTGATAGCTTAAAGAAACATTCGCTACAGGCTCGGAGTA 1443
495 AsnIleSerSerIleAspGlyAlaLysLysAlaLys...IleGluThrLys 510
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1444 GATATGATAGCTTAAGCTCATCTCAACAGCAGCTGTTATTAACACAA 1493
510 sAlaIleThrSerLysAsnLeuThrLeuSerGlyThrIleThrLeuLeuAsp 527
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1494 CACCGCAAAATTAACAGATATCCGTAGCGACTATAGAATTTATCTTCC 1543
527 rThrGlyThrPheThrGluAsnHisSerLeuArgAsnProGlnSerThr 543
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1544 CTACTGCGAATGGCTTGAAGATCTCAGAAATGGAATTCACAGAGCTTC 1593
544 AspIleLeuGlnLeuLys.....AlaSerGlyThrValThrSerThrAl 558
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1594 CCTCTGCTCTTAAAGCCTGAGCCGGGGAGTACGTAAGTACTGC 1643
558 a.....ValThrProAspProIleMetGlyGlyLysPheH 570
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1644 TGGAGATTTCTTACCGGTAAAGTCC.....C 1669
570 lAspGlyThrGlnGlyThrThrGlyProIleValThr...GlyThrGly 585
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1670 ATATGCTTTTCAAGCAATGG...AAATGAGCTTGCAGAGAGACTGGA 1716
586 AlaSerThrThrAlaThrPheAsnThrThrLysThrGlyThrIleProAs 602
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1717 ...AACAAAGTTGGAATTTCTCTGGATAAATAATTAATTAAGCTAG 1763

```


Align seg 1/1 to: A27342 from: 1 to: 3150

```

1 MetLysSerSerPheProLysPheValPheSerThr..... 12
  |||||.....
101 ATGAAACGCTATTCGTAGTCTTAATTCTTACACACTGGCGGCATG 150
  |||||
13 PheAlaIlePheProLeuSerMetIleAlaThrGluThrValLeuAspS 29
  |||||
151 TTTTGCT.....TCACACACGCTTCTACTAGAGTATCATCTC 188
  |||||
29 eSerAlaSerPheAspGlyAsnLysAsnGlyAsnSerValArgLys 45
  |||||
189 CTTCGAGAGACTTGATGATCGAGGAGGAGATTTTCTCTACACAA 238
  |||||
46 SerGlnGluAspAlaGlyThrThrTyrLeuPheLysGlyAsnValThrLe 62
  |||||
239 CTTCGTGATCCTAGAGGACACTCTGATTTTTCAGGGATCTCTACAT 288
  |||||
62 uGluAsnIleProGlyThrGlyThrAlaIleThrLysSerCysPheAsnA 79
  |||||
289 TCCGAACTCTGATTAATGCGCATATCCAGAACCTCTCCAGTTCCTTA 338
  |||||
79 sPThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuPhe 95
  |||||
339 ATAGGGCGGAGACACTACAATCTTAGAAGAGTGGGCTTCTCTCCTC 388
  |||||
96 GlnThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSerSerVa 112
  |||||
389 TTTAAATATC...CGTTCTCAGCTGACGGAGCCGCGATTAGTAGTAT 435
  |||||
112 LValAspLys.....SerThrPheIleGlyPheSerSerL 125
  |||||
436 CACCCAAAATCTCTGAACATATGTCCTTGAGTTTTCAGGATTTAGTCA 485
  |||||
135 eUSeSerPhe.....IleAlaSerProLysSerSer 134
  |||||
486 TGATCTTCGATTAAGTGAATCTTGACTTCAGATACCTCAGCGAGTAT 535
  |||||
135 IleThrThrGlyLysGlyAlaValAsnSerSerThrGlySerLeuSerie 151
  |||||
536 GTCATACCTCAGCAGCATCGCGATTACGCTACACGCCCATGCTC...TT 582
  |||||
151 uThrLysAsnValSerLeuLeuPheSerLysAsnPheSerThrAspArg 168
  |||||
583 TACAACAATGACTCATCTCATCTTCCAATACACCGTTCTCGACGATTGG 632
  |||||
168 LysIleAlaIleThrAlaLysThrLeuSerSerLeuThrGlyThrThrMetSer 184
  |||||
633 GAGCTGCATTCGAGGACACACATCACAATAGAAAATACGAAAAAGAC 682
  |||||
185 AlaLeuPheSerGluAsnThrSerSerLysGlyGlyAlaIleGlnThr 201
  |||||
683 CTTCCTTTAATGTAATGATGATCTCTAATGAGGGGCCCTCAGCGG 732
  |||||
201 rSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerPheSerVa 218
  |||||
733 ATCTCAGAGATCAACCTCATCAACATAGCCCTCTGATTTCTCA 782
  |||||
218 sp..... 218
783 CGAATGTACAGAGATCTATGTGGGCTATTACCTACCGAGAGATCT 832
  |||||
219 .....AsnThrSerSe 222
  |||||
833 ATGCTACACCTCTGGAGACCTCTCAGAGAGCTTGTTGCTTAATATAGCTC 882
  |||||
222 rAspSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAspVa 239
  |||||
883 GCGCTCAGAGAGCGCATCTATGCTAAGCAATATGCACATTTTCTAA 932
  |||||
239 snAlaLysValSerPheIleAspAsnLysValThrGlyAlaSerSer... 254
  |||||

```

```

933 ACAGCGACCTGACTTCCAAAACATACAGCATCTCCACAAAACCTCTTA 982
  |||||
255 .....SerThrThrGlyAspMetSe 261
  |||||
983 CCGACCTACACCTCCACCTACACACAGAGAGTCACTCTGTTGTAGG 1032
  |||||
261 r...GlyGlyAlaIle...Cys.....AlaTyrLysThrSerThr 273
  |||||
1033 ATATGAGCGCCCATCTTCTGACTCTCCAGCTACCCGCCACCAACAG 1082
  |||||
273 sPThrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAsn 289
  |||||
1083 GTCTTAGCTGACTATATCTGGAGAAACAGCGTACTTCTTCAAAAC 1132
  |||||
290 ThrSerThrThrAlaGlyAlaIleTyrValLysLysLeuGluLeu 306
  |||||
1133 ATTGCTCCGAAACAGAGAGAGCCCTCTATGCAAAAGATCTATAG 1182
  |||||
306 aSerGlyLeuThrLeuPheSerArgAsnSerValAsnGlyGlyThr 323
  |||||
1183 TTCTATATAATCTACAAATATTTCTT.....GGAATACAG 1217
  |||||
323 LaProLysGlyAlaAlaIleAlaIleGluAspSerGlyLysLeuSerLeu 339
  |||||
1218 CTGAAAGAGAGCGCCATGCTATGCTATCCGAATCTGGGAGACTCTCT 1267
  |||||
340 SerAlaAspSerGlyAspIleValPheLeuGlyAsn...ThrValThrSe 355
  |||||
1268 TCCGCAATCAAGGTGATATCTCTTTAACAAAGACCTCAGATCAG 1317
  |||||
355 ThrThrProGlyThrAsnArgSerSerIleAspLeuGlyThrSerAla 372
  |||||
1318 TGGGACACT....ACTCGCAATAGATTTCACTCGGAAAGATGCCA 1361
  |||||
372 ySmeThrAlaLeuArgSerAlaAlaGlyArgAlaIleTyrPheTyrAsp 388
  |||||
1362 AGTTGGCAGCTCTAGGAATAGCGAAGCTATACCTTACTTCTATGAT 1411
  |||||
389 ProIleThrThrGly.....SerSerThrThrValThrAspValLeu 403
  |||||
1412 CCGATTACATGATGATTAATCTGCTGCATCCGAGCGCTACTGCTGT 1461
  |||||
403 sValAsn...GluThrProAlaAspSerAlaLeuGlnThrThrGlyAsn 419
  |||||
1462 CGTCAATCCCAAGCAGCTCGAGATGTCG....TATTCAGGAGCTA 1505
  |||||
419 IleLeuPheThrGlyLysLeuSerGluThrGluAlaAlaAspSerLys 435
  |||||
1506 TTGCTCTTTTCAGAGAAACCTCTCACTGCTACCGAAGCAGCAACCCCTCA 1555
  |||||
436 AsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGlyGlyThrLe 452
  |||||
1556 AATGCTACATCTACTATTAACCAAAAGCTAGAACTGAAAGCGGACTCT 1605
  |||||
452 uSerLeuLysHisGlyValThrLeuGlnThrGluAlaPheThrGlnGln 469
  |||||
1606 CGCTTAAAGAAAGCGTGTACTTAAATGTCATACACTTCACGCAAGTGC 1655
  |||||
469 LaAspSerArgLeuGlnMetAspValGlyThrThrLeuGluProAlaAsp 485
  |||||
1656 AAAAGTCCGCTGTCATCATGATGAGGAGCACATAGCAACTACAAAT 1705
  |||||
486 ThrSer.....ThrIleAsnLeuValIleAs 495
  |||||
1706 GAGCTAATTAATCTGACGCGTCTATCACTTAACAAAGCTTGTAATCA 1755
  |||||
495 nIleSerSerIleAspGlyAlaLysLysAlaLysIleGluThrLysAla 512
  |||||
1756 TGTGATTCCTTGGATGAGCTAAAGCGCTGCTTAATGTCAGAGCTA 1805
  |||||
512 hrSerLysAsnLeuThrLeuSerGlyThrIleThrLeuLeuAspProThr 528
  |||||
1806 CCAATGAGCTCTCATATATCCGAACTTGAAGACTTGTGAAAAAAGCTCT 1855
  |||||

```

539 GlyThrPheTyrGluAsnHisSerLeuArgPro.....GlnSerTyr 543
:::|||||::: |||
1856 CAAAGATTGGTGTGACACACCGGATGTTAATAAGATTACACAGAGT 1905
543 rasPleuGluLeuLysAla...SerGlyThrValThrSerThrAlaVal 559
|||||::: |||
1906 TCCGATTTCAGAACTCAAGCGACTCAAAATGTAAGTAACTACAGGACT 1955
559 alThrProAspProIleMetGlyIuLysPheHisTyrGlyTyrGlnGly 575
::: |||
1956 TCAGTCTGGCCACAACGGCTATCAGCAATCTCCCTATGGGTATCAAGA 2005
576 ThrTrpGlyProIleValTrpGlyThrGlyAlaSerThrThrAlaThrPh 592
|||||::: |||
2006 ACTTGGAGATTACCATAGACACGACA.....ACCCATACGGTCAACAG 2049
592 eAsnTrpThrLysThrGlyTyrIleProAsnProGluArgIleGlySerL 609
|||||::: |||
2050 AAATTGGAAAAAACCGGTTATCTTCATCCGAGCGCTTCTGCCCTCCC 2099
609 euValProAsnSerLeuThrPAsnAlaPheIleAspIleSerSerLeuHis 625
|||||::: |||
2100 TCATTCCCTAATAGCCTATGGGCAACGTCATAGATTTCAGAGCTGTAGT 2149
626 TyrIleuMetGluThrAlaAsnGlnGlyLeuGlnGlyAspArgAlaPheTr 642
::: |||
2150 CAAGCGTCAGCAGCTGATGGCAAGATGTCCCTGGG...AAGCACTAG 2196
642 PCysAlaGlyLeuSerAsnPhePheHisLysAspSerThrLysThrArg 659
::: |||
2197 CATCACAGAAATTACAAATTTCTTCATGCAATCATACCGGTGATGCAC 2246
659 rGlyLysPheArgHisLeuSerGlyGlyTyrValIleGlyLysLeuHis 675
|||||::: |||
2247 GCAGGTACCGCATATGGGTGAGGCTCATCATCAATACCTACACAGCG 2296
676 ThrCysSerAspLysIleLeuSerAlaAlaPheCysGlnLeuPheGlyAr 692
::: |||
2297 ATCACTCCAGATGCTGCGTTAAGTCTGAGTTTGGACAGCTGTTACAA 2346
692 gAspArgAspTyrPheValAlaLysAsnGlnGlyThrValTyrGlyTyr 709
::: |||
2347 ATCTAAGGATTAAGCTCTGAGTACAGGTCATCTTAACGTTATTCGCTA 2396
709 hrLeuTyrTyrGlnHisAsnGluThrTyrIleSerLeuProCysLysLeu 725
::: |||
2397 CAGTATACTCTAATCATCACCAAGTCTCTGTTTGA..... 2431
726 ArgProCysSerLeuSerTyrValProThrGluIleProValLeuPheSe 742
::: |||
2432TCATCGAGATTCTTCTC 2448
742 rGlyAsnLeuSer.....TyrThrHisThrAspAsnAspLeuLysT 756
|||||::: |||
2449 AGGAGGCACTCTCGAGTACCTATACCCGACATGAGAAAGTAAAGA 2498
756 hrLysTyrThrThrTyrProThrValLysGlySerTrpGlyAsnAspSer 772
|||||::: |||
2499 CTTCATATACAAATTCCTAAAGGCGCTGCTCTTGAGTAACAATTGC 2548
773 PheAlaLeuGluPheGlyArgAlaProIleCysLeuAspGluSerAl 789
::: |||
2549 TGGTTAGGAGACTCGAAGGAACTTCCTCATCTCTCTTCGCAAT 2598
789 aleu...PheGluGlnTyrMetProPheMetLysLeuGlnPheValTyr 805
|||||::: |||
2599 CTTAACCTCAAGCAGATCATTCCTTTGTAAGAAGTGAAGTTCTTACG 2648
805 lahSgInGlnGlyPheLysGlnGlnGlyThrGluAlaArgGluPheGly 821
||::: |||
2649 CGACTCATGGGGCATCCAGAAATATACCCCGAGGAGGATTTTGTGA 2698

822 SerSerArgLeuValAsnLeuAlaLeuProIleGlyTyrLeuArgPheAspLys 838
::: |||
2699 CACGGTCATCTACTCAACGTTCGAGTTCCCGTAGGCGTCCGCTTGGTTAA 2748
838 sGluSerAspCysGlnAspAlaThrTyrAsnLeuThrLeuGlyTyrThrV 855
::: |||
2749 AAATTTCATATATGACACAGATTTTTCACATAATATGTAAGCTATGCTC 2798
855 alaPheLeuValArgSerAsnProAspCysThrThrThrLeuArgIleSer 871
|||||::: |||
2799 CTGATGCTATCGTCACAAATCCGATTCGATTCGATACGACATTAATTAAT 2848
872 GlyAspSerTrpLysThrPheGlyThrAsnLeuAlaArgGlnAlaLeuVal 888
||::: |||
2849 GGACTACGTGAGCTCTATAGGAAATATCTAACCAAGAGTACTTGGCT 2898
888 lleuArgAlaGlyAsnHisPheCysPheAsnSerAsnPheGluAlaPheS 905
|||||::: |||
2899 AGTACAGACATCCAGCCATCTTCAATATGATGTTCTAGACATCTTGC 2948
905 ergInPheSerPheGluLeuArgGlySerSerArgAsnTyrAsnValAsp 921
::: |||
2949 GGCACGTGTGATGATATTTCGAGAACCTCCGTAATAATTAATCTAGAT 2998
922 leuGlyAlaLysTyrGlnPhe 928
::: |||
2999 ATAGGAAGCAAAATTACGATTT 3019

Fri Nov 24 13:50:05 2000

us-09-428-122-2.rng

Page 33


```

56 PheLysGlyAsnValThrLeuGluAsnIleProGlyThrGlyThrAlaIle 72
   ::::::::::::::::::::
2105 ATAAAGCAATATATCC.....AACAAATCAGTGATTTGAATATATAT 2148
72 eThrLysSerCys.....PheAsnAsnT 80
   ::::::::::::::::::::
2149 TGAATAAAAGGCAAGCGTGAATCCAAATGGCGCAATATATCTCACAA 2198
80 hTrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPheGln 96
   ::::::::::::::::::::
2199 AAGAGGCAATCTCACAATTTCTTCTGATTAAGTAATATATACCAATCAG 2248
97 ...ThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSerSerVa 112
   ::::::::::::::::::::
2249 ATACCAATCAAAAGCAGCGTTGAAGGGGCGTTGATTCAGTACGTAGCG 2298
112 lValAspLysSerThrThr...../..... 118
   ::::::::::::::::::::
2299 AGAAATGCTAACCTTACTTCAAAACCAAGTTAAATTGGCAGAG 2348
119 .....PheIleGlyPheSerSerLeuSerPheIleAlaSerProGly 132
   ::::::::::::::::::::
2349 ACCTAAATATTTTCAGGCTTAAATAAGCAGAAATACAGCTAAATATGCG 2398
133 SerSerIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerle 149
   ::::::::::::::::::::
2399 AGTGATTTACTATTGGC.....AATGCTAGCGGTGTAATGTC 2436
149 uSerLeuThrLys..... 153
   ::::::::::::::::::::
2437 TGATGCTAAAAAGTGACTTTGACAAAGTTAAAGTTCAAAATCTCGA 2486
154 .....AsnValSerLeuLeuPheSerSerLeuAsnPheSerThrAsp 166
   ::::::::::::::::::::
2487 CTGACGCTACAAATGTAACACTA.....AATAGCGAAGTGAAACGTCCT 2530
167 AsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThMe 183
   ::::::::::::::::::::
2531 AATGTAGTAGGC...AATGCTGTATATGATACACACCGGTTTATACCAT 2577
183 tSerAla.....LeuPheSerGluAsnThrSerSerLysLys.... 195
   ::::::::::::::::::::
2578 TTCCCGCAAAAGATGTACGTAACAAATACGTTACCTCCCAACAAAGCA 2627
196 .....GlyGlyAlaIleGlnThrSerAspAlaLeuThr 206
   ::::::::::::::::::::
2628 TAAATATCTCTGCCGACAGCAAGAAATGTAACACCAAGAGGCACTACT 2677
207 lIleThrGlyAsnGlnGlyGlyValSerPheSerAspAsnThrSerSerAs 223
   ::::::::::::::::::::
2678 ATCAATGCAACCAAGCAGCGGTGAGTAAGTACTGCTCAAAATGCTACAT 2727
223 pSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsnA 240
   ::::::::::::::::::::
2728 TAAAGGCAC.....ATTACCTGCAAAATGTAACAGTACAGCAACAG 2771
240 lAlysValSerPheIleAspAsnLysValThrGlyAlaSerSerSerThr 256
   ::::::::::::::::::::
2772 AAAATCTGTGTACACAGAGATGCTCATTAATGCAACCAAGCGCAC 2821
257 .....ThrGlyAspMetSerGlyGlyAlaIleCysAl 267
   ::::::::::::::::::::
2822 GTAACATGTAGTACAAAACAGGGGATATTAAAGTGGA.....AT 2862
267 aTYrLYSThSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
   ::::::::::::::::::::
2863 TGAATCAACTTCGGTAATGTAATATTAACAGGACGCGCAAT..ACAC 2909
284 euLeuPheSerAsnAsnThr..... 290
   ::::::::::::::::::::
2910 TTAAGTAGTAATATACCTGCTCAAGATGTAAACAGTAAACAGCGATGCA 2959
291 .....SerThrThrAlaGlyAlaIleTYrValLys..... 301

```

```

2960 GGAGCGCTGACACACTACAGCAGCGCTCAACCAATTATGTCGACAAAGCGCA 3009
   ::::::::::::::::::::
3002 .....LysLeuGluLeuA 306
   ::::::::::::::::::::
3010 TGCAAAATATTACAAACCAAAACAGGTGATATCAACGGTAAGTTGAATCCA 3059
306 lAspGlyGlyLeuThrLeuPheSerIArgAsnSerValAsnGlyGlyThr 322
   ::::::::::::::::::::
3060 GCTCCGGCTCTGTACACTTGTGTGCAACTGGAGCAACTCTGCTGTAGGT 3109
323 AlaprolYsglyGlyAlaIleAlaIle...GluAspSerGlyGluLeu.. 337
   ::::::::::::::::::::
3110 AATATTTCAGGTACACTGTTACTATTACTGCGGATAGCGGTAATATTAC 3159
338 .....SerLeuSerA 341
   ::::::::::::::::::::
3160 CTCACACGTAGGTTCTACAAATTAATGGAGCTAATAGTATACACCTCA 3209
341 lAspSerGlyAspIle.....ValPheLeuGlyAsnThrValThrSer 355
   ::::::::::::::::::::
3210 GCCAATACGCGCATATTGAAGTAGACAAATTCGTGTAATACACTAATGTT 3259
356 ThrThrProGlyThrAsnArgSerSerIleAspLeuGlyThrSerAlaLy 372
   ::::::::::::::::::::
3260 ACA.....GCAAGCACTGGTGATTTAACTATTGGAATAAGTGCMAA 3300
372 sMetThrAlaLeuArgSerAlaAlaGlyArgAlaIleTYrPheTYrAsp 389
   ::::::::::::::::::::
3301 AGTTGAAGCGAAATGAGAGCTGCA..... 3325
389 rolIeThrThrGlySerSerThrThrValThrAspValLeuLysValAsn 405
   ::::::::::::::::::::
3325 ..... 3325
406 GluThrProAlaAspSerAlaLeuGlnThrThrGlyAsnIleIlePheTh 422
   ::::::::::::::::::::
3325 ..... 3325
422 rGlyGlyLysLeuSerGlyThrGluAlaAlaAspSerLysAsnLeuThrs 439
   ::::::::::::::::::::
3326 .....ACCTTAAGCTGCAATCGACCAAAATTAACCA 3356
439 eRlyLeuLeuGlnProValThrLeuSerGlyGly.....ThrLeuSer 453
   ::::::::::::::::::::
3357 CCCAACAAGCGCTACGATTAATCTCAACGATGTGTACACACTCTTACA 3406
454 LeuLysHis.....GlyValThrLe 460
   ::::::::::::::::::::
3407 GCCAAGATAGCACTATCGCAAGAAATTAATGCTGTAATGTGAGCTT 3456
460 uGlnThrGln...AlaPheThrGlnGlnAlaAspSerArgLeuGluMetA 476
   ::::::::::::::::::::
3457 AAATACCAAGCGCACTTAATCTACTACAGGGGATTCAAAGATTAAAGC 3506
476 sPVAlGlyThr...ThrLeuGluProAlaAsp.....Thr 486
   ::::::::::::::::::::
3507 CCAGTGTACTTAACTAATCAATGCAAAAGATGCCAATTAGTGTGCT 3556
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLy 503
   ::::::::::::::::::::
3557 GCATCAGTAGTACCGCACAGTAGTAATGCAACTAACGCAAGGCTCTGG 3606
503 sLYsAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLysusG 520
   ::::::::::::::::::::
3607 TAACGTG.....ACTGCGAAACCTCAAGCAGCGGGAATATACACCG 3647
520 lYThrIleThrLeuLeuAspProThrGlyThrPheTYrGluAsn 534
   ::::::::::::::::::::
3648 GGGATTTAAACACAAATTAATGGTTAAATATATCATTTCCGAAAT 3691
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-302-832-7

```



```

seq_documentation_block:
: Sequence 7, Application US/08302832
: Patent No. 5603938
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matzare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/302,832
: FILING DATE: 16-SEP-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US pct/us93/02166
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-404
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4287 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-302-832-7

alignment_scores:
: Quality: 271.50 Length: 648
: Ratio: 0.828 Gaps: 29
: Percent Similarity: 50.617 Percent Identity: 22.531

alignment_block:
US-09-428-122-2 x US-08-302-832-7 ..
Align seq 1/1 to: US-08-302-832-7 from: 1 to: 4287

23 ThrGluThrValLeuAspSerSerAlaSerPheAspGlyValnLysAsnG1 39
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1955 ACAATATACCTTTTATGTAGCGCGCTCATTTGACACAAATGCGCGTTC 2004
39 yAsnPheSerValargLuser..... 46
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2005 AACATTTCATTGCCAGAGAGGCGCTAAATTAAGATATCATATACA 2054
47 .....GlnGluAspAlaGlyThrThrTyr.....Leu 55
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2055 CCAGTACCTTAATATTACCAACCACTGTATACCACTTACCGCACCAT 2104
56 PheLysGlyValnValThrLeuGluValnLeuProGlyThrGlyThrAlaI1 72
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2105 ATAAAGGCAATATATATCC.....AACAAATCAGGTGATTGAAATATAT 2148

```

```

72 eThrLysSerCys.....PheAsnAsnT 80
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2149 TGATTAATAAAGGACGCTGAATTCCAATTTGGCGCAATATCTCCAA 2198
80 hTrysGlyAspLeuThrPheThrGlyAsnGlyValnSerLeuLeuPheGln 96
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2199 AAGAAGCAATCTCACATTTCTTCTGATTAAGTAATATTACCAATCAG 2248
97 ...ThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSerSerya 112
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2249 ATAAATCAATCAAGCAGCGCTGAAGGGGCGTTCGATTCAAGTAGCG 2298
112 lValAspLysSerThr..... 118
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2299 AGAAATGCTAACCTAATTCAACCAAGAGTTAAATTTGGCAGAG 2348
119 .....PheIleGlyPheSerSerLeuSerPheIleAlaSerProGly 132
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2349 ACCTAATATTTCAGCGCTTAATAAAGCAGAAATTCACCTAATAATGCG 2398
133 SerSerIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerle 149
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2399 AGTGAATTAACTATTGCG.....AATGCTAAGCGGTGTAATGCG 2436
149 uSerLeuThrLys..... 153
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2437 TGATGCTAATAAAGTACTTTTGACAGGTTAAAGATTCAAAAATCTCGA 2486
154 .....AsnValSerLeuLeuPheSerLysAsnPheSerThrAsp 166
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2487 CTGACGCTCAATGTAACACTA.....AATAGCAAGTGAACGCT 2530
167 AsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThme 183
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2531 AATGTAAGTAGC...AATGCTGATGATTAACACACCGCTTTAACCAT 2577
183 tSerAla.....LeuPheSerGluAsnThrSerSerLysLys.... 195
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2578 TTCGGCAAAAGATGTAACGTTAAACATATACCTTCCCAAGACAA 2627
196 .....GlyGlyAlaIleGlnThrSerAspAlaLeuThr 206
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2628 TAAATATCTCTCCGCGCAGCAAGTAATGTAAACCAAGAAAGCACACT 2677
207 lIleThrGlyAsnGlnGlyValAlaSerPheSerAspAsnThrSerSera 223
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2678 ATCAATGCAACCAAGCAGCGTGAAGTAACTGCTCAAAATGTACAT 2727
223 pSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsna 240
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2728 TAAAGGCAC.....ATTACTCGCAAAATGTACAGAGACAGACAG 2771
240 lAlaValSerPheIleAspAsnLysValThrGlyAlaSerSerThr 256
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2772 AAAATCTTGTACACAGAGAAATGCTCATTAATGCAACGCGCAGCA 2821
257 .....ThrGlyAspMetSerGlyValAlaIleCysAl 267
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2822 GTAACATTAGTACAAAAACAGGGATATTAAGGTGA.....AT 2862
267 aThrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2863 TGAATCAACTCCGTAATGTAAATATTAACGCGAGCGCAAT...ACAC 2909
284 eLeuPheSerAsnAsnThr..... 290
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2910 TTAAGTAGATATATCACTGCTCAAGATGTAACAGTAACAGCGGATGCA 2959
291 .....SerThrThrAlaGlyGlyAlaIleLysValLys..... 301
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2960 GGAGCGTTGACAACTACAGCAGCGCTCAACATTAAGTGGACACAGCGCA 3009

```

```

302 .....LysLeuGluLeuA 306
3010 TCGAATATTACAAACCAAGCATATCAACGGTAAGTTGATCCA 3059
306 1aSerGlyLeuThrLeuPheSerArgAsnSerValasnGlyThr 322
3060 GCTCCGCGCTGTGAACACTGTGGCAACGAGCAACTCTGCTAGT 3109
323 AlaProGlyGlyAlaIleAlaIle...GluAspSerGlyGluLeu.. 337
3110 AATATTTCAGTAACACTGTACTTACTTACTGCGGATAGCGTAATTAAC 3159
338 .....SerLeuSerA 341
3160 CTCACAGTAGGTTCTACATTAATGAGCTAATAGTGTACACCACTCA 3209
341 1aAspSerGlyAspIle.....ValPheLeuGlyAsnThrValThrSer 355
3210 GCCAATCGCGCATATTGAGGTACATTTCTGTAATACGTAATAGTT 3259
356 ThrThrProGlyThrAsnArgSerSerIleAspLeuGlyThrSerAlaIly 372
3260 ACA.....GCACGCACTGGTATTAAGTACTTATGGAATAGTGA 3300
372 SmetThrAlaLeuArgSerAlaIleAlaGlyArgAlaIleTyrPheTyrAsp 389
3301 AGTGAACGCAAAATGAGCTGCA..... 3325
389 rollerThrGlySerSerThrThrValThrAspValLeuLysValAsn 405
3325 ..... 3325
406 GluThrProAlaAspSerAlaLeuGlnTyrThrGlyAsnIleIlePheTh 422
3325 ..... 3325
422 rGlyGluLysLeuSerGlyThrGluAlaAlaAspSerLysAsnLeuThrs 439
3326 .....ACCTTAAGTCTGTAATCAAGCAAAATTAAACA 3356
439 eRlyLeuLeuGlnProValThrLeuSerGlyGly.....ThrLeuSer 453
3357 CCCAACAAGCGCTTACATTACTCAAGCAATGTCAGCACTTACCA 3406
454 LeuLysHis.....GlyValThrIle 460
3407 GCCAAGATAGCAGTATCCGAGAAACATTAATGCTTAATGTGACGTT 3456
460 uGlnThrGln..AlaPheThrGlnGlnAlaAspSerArgLeuGluMetA 476
3457 AATATCCAGAGCGACTTAAGTACTACAGGAGGATCAAGATTAAACGA 3506
476 sPValGlyThr...ThrLeuGluProAlaAsp.....Thr 486
3507 CCAGTGTGACCTTAACATCAATGCAAAAGATGCCAAATTAGATGTGCT 3556
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaIly 503
3557 GCATCAGTGTGACCGACAGTACTAATGCAACTAAGCAAGGCGTCTGG 3606
503 slyAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerG 520
3607 TAACTG.....ACTGGCAAAACCTCAAGCAGCGTGAATATCAACG 3647
520 1yThrIleThrLeuLeuAspProThrGlyThrPheTyrGluAsn 534
3648 GGGATTTAACACATAAATGGTTAAATATCATTTTCGAAAT 3691
seq_name: /cgn2_6/prodata/2/ina/5C_COMB.seq:us-08-530-198-7
seq_documentation_block:
; Sequence 7, Application US/08530198
; Patent No. 5869065

```

```

? GENERAL INFORMATION:
? APPLICANT: BARENKAMP, STEPHEN J
? APPLICANT: ST. GEME LTD, JOSEPH W
? TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
? TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESS: Shoemaker and Mattare, Ltd
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22202-0286
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/530,198
? FILING DATE: 13-DEC-1995
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: BERKSTRESSER, JERRY W
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: JWB-1186
? TELEPHONE: (703) 415-0813
? TELEFAX: (703) 415-0810
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4287 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? US-08-530-198-7

alignment_scores:
Quality: 271.50 Length: 648
Ratio: 0.828 Gaps: 29
Percent Similarity: 50.617 Percent Identity: 22.531

alignment_block:
US-09-428-122-2 x US-08-530-198-7 ..

Align seg 1/1 to: US-08-530-198-7 from: 1 to: 4287

23 ThrGluThrValLeuAspSerSerAlaSerPheAspGlyAsnLysAsnG1 39
1955 ACAATTAACACTTTAATGTAAGCGGCTCATTTGACACAAATGCGCTTC 2004
39 yAsnPheSerValArgLysSer..... 46
2005 AAACATTTCATTCGACAGAGAGGGCTAAATTTAAAGATATCATATACA 2054
47 .....GlnGluAspAlaGlyThrThrTyr.....Leu 55
2055 CCAGTAGCTTAATATATCCACCACCACTGATACCATTTACCGCACCAAT 2104
56 PheLysGlyAsnValThrLeuGluAsnIleProGlyThrGlyThrAlaI1 72
2105 ATAAAAAGCAATATATCC.....AACAAATCAGGTGATTGAAATATTAT 2148
72 eThrLysSerGys.....PheAsnAsnT 80
2149 TGATAAAAAAGCGACGCTGAATCCAAATTTGGCGGCAATATCTCACAAA 2198
80 hLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPheGln 96
2199 AAGAAGCAATCTCACAAATTTCTTGTGATAAGTAATATTTACCAATCAG 2248

```



```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS.vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-7

```

alignment_scores:

Quality:	271.50	Length:	648
Ratio:	0.828	Gaps:	29
Percent Similarity:	50.617	Percent Identity:	22.531

alignment_block:

US-09-428-122-2 x US-08-469-880-7 ..

Align seg 1/1 to: US-08-469-880-7 from: 1 to: 4287

```

23 ThrGluThValIeuAspSerSerAlaSerPheAspGlyAsnLysAsnG1 39
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
1955 ACAAAATTACCTTTAATGTACGCCGCTCATTTGACAAACAATGCGCTTC 2004
39 yAsnPheSerValArgGluSer..... 46
2005 AAACATTTTCATTGCCAAGAGGGGCTAAATTAAAGATATCAATAACA 2054
47 .....GlnGluAspAlaGlyThrThrTyr.....Leu 55
2055 CCAGTAGCTTAATATATACCAACAACCTGTATACCACTTACCGCACTT 2104
56 PheLysGlyAsnValThrLeuGluAsnIleProGlyThrGlyThrAlaI1 72
||||| :||||| :||||| :||||| :||||| :||||| :||||| :
2105 ATAAAGGCAATATATATCC.....AACAAATCAGGTGATTGATATATAT 2148
72 eThrLysSerCys.....PheAsnAsnT 80
||||| :||||| :||||| :||||| :||||| :||||| :||||| :
2149 TGATATAAAAGGCAAGCGCTGAATCCAAATTGGCGGCAATATATCCAAA 2198
80 hrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPheGln 96
||||| :||||| :||||| :||||| :||||| :||||| :||||| :
2199 AAGAGGCAATCTCACAATTTCTCTGATTAAGTAATATATACCAATCAG 2248

```

```

97 ...ThrValAspAlaGlyThrValAlaGlyAlaValAsnSerSerVal 112
||||| :||||| :||||| :||||| :||||| :||||| :||||| :
2249 ATACAAATCAAAACAGCGCTTGAGGGGGCGTCTGATTAAGTAGAGCG 2298
112 ValAspLysSerThrThr..... 118
2299 AGAAATGCTAACCTACTATTCACCAACCAAGAGTTAAATTCGACAGAG 2348
119 .....PheIleGlyPheSerSerLeuSerPheIleAlaSerProGly 132
2349 ACCTAAATATTTCAGCGCTTTAATAAGCAAGAAATATACAGCTTAAATG 2398
133 SerSerIleThrThrGlyLysGlyValAlaLysCysSerThrThrGly 149
2399 AGTAGATTATACATATGCG.....AATGCTAGCGGTGGTAAATGCG 2436
149 uSerLeuThrLys..... 153
2437 TGATGCTAAATAAGTAGCTTTTGACAAAGTTAAAGATTCAAAATCTCGA 2486
154 .....AsnValSerLeuLeuPheSerLysAsnPheSerThrAsp 166
2487 CTGAGCGTCACAAATGTACACTA.....AATAGCGAAGTGAAACCTCT 2530
167 AsnGlyLysAlaIleThrAlaLysThrLeuSerLeuThrGlyThrPhe 183
||||| :||||| :||||| :||||| :||||| :||||| :
2531 AATGTTACTAGC...AATGCTGTAAATGATATACAGCCGCTTAAACAT 2577
183 tSerAla.....LeuPheSerGluAsnThrSerSerLysLys.... 195
2578 TTCGCGCAAAAGATGTAACGGTAAACAAATAAGCTTACCTCCCAAGACAA 2627
196 .....GlyGlyAlaIleGlnThrSerSerAlaLeuThr 206
2628 TAAATATCTCTGCCCGCAGAGAAATGTACACCAAAAGGACACACT 2677
207 IleThrGlyAsnGlnGlyGluValSerPheSerAspAsnThrSerSer 223
||||| :||||| :||||| :||||| :||||| :||||| :
2678 ATCAATGCAACACAGCGCAGCGGAGTAAGTACGCTCAAAATGTATCAAT 2727
223 pSerGlyAlaIlePheThrGluAlaSerValThrIleSerAsnAsn 240
||||| :||||| :||||| :||||| :||||| :||||| :
2728 TAAAGCGAAC.....ATTACCTCGCAAAATGTACAGTACGACGACACAG 2771
240 LysValSerPheIleAspAsnLysValThrGlyLysSerSerThr 256
2772 AAATCTTGTACACAGAGAAATGCTGATTAATGCAACGCGCGCACA 2821
257 .....ThrGlyAspMetSerGlyGlyAlaIleCysAl 267
2822 GTAACATTATGTAACAAACAGGAGATATTAAGTGA.....AT 2862
267 aTryLysThrSerThrAspThrLysValThrLeuThrGlyAsnIle 284
2863 TGATATCACTTCGGTAAATGTAATATATACAGCGAGCGCAAT...ACAC 2909
284 euleuPheSerAsnAsnThr..... 290
||||| :||||| :||||| :||||| :||||| :||||| :
2910 TTAAGTAGTAATATCACTGCTCAAGATGTACACAGTACAGCGGATGCA 2959
291 .....SerThrThrAlaGlyAlaIleThrValLys..... 301
2960 GGAGCGCTTGACAACTACAGCAGCTCAACCTTGTGCGAACACAGCGCA 3009
302 .....LysLeuGluLeuA 306
3010 TGCAAAATTTCAACCAACAGGTGATATCAACGGTAAAGTTAGATCCA 3059
306 lAserGlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyGlyThr 322
||||| :||||| :||||| :||||| :||||| :||||| :
3060 GCTCGCGCTGTGAACACTTGTGCACTGAGACACTCTGCTGTAAGT 3109

```

```

333 AlaProlysglyAlaIleAlaIle...GluAspSerGlyGluLeu.. 337
      |||
3310 AATATTGAGTAACACTGTACTATCTACTGCGATACGGTAATATAC 3159
      |||
338 .....SerLeuSera 341
3160 CTCACAGTACAGTCTACAAATTAATGAGCACTAATAGTACACCACTCA 3209
341 LaasPserGlyAspIle.....ValPheLeuGlyAsnThrValThrSer 355
      |||
3210 GCCAATCAGCGGATATGAGGTACATTTCTGTAATACAGTAATATGTT 3259
356 ThrThrProGlyThrAsnArgSerSerIleAspLeuGlyThrSerAlaLy 372
      |||
3260 ACA.....GCAAGCAGCTGCTGATTTACTATTGAGAAATAGTCGAA 3300
372 SmeThrAlaLeuArgSerAlaIleGlyArgAlaIleThrPheThrAsp 389
      |||
3301 AGTTGACGCAAAATGAGCTGCA..... 3325
389 rollerThrThrGlySerSerThrThrValThrAspAlaLeuValAsn 405
3325 ..... 3325
406 GluThrProAlaAspSerAlaLeuGlnThrThrGlyAsnIleIlePheTh 422
3325 ..... 3325
422 rGlyGluLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeuThrs 439
      |||
3326 .....ACCTTACCTGCTGATCAGCAATCAATTAACCA 3356
439 eRysLeuLeuGlnProValThrLeuSerGlyGly.....ThrLeuSer 453
      |||
3357 CCCAAACAGCGCTAGCATTTACTTACCAACAGTGTACAGACAACTCTTACA 3406
454 LeuYsHis.....GlyValIle 460
3407 GCCAAGGATAGCATGTGCGAGAAACATTAACTGCTTAATGTGACGTT 3456
460 uGlnThrGln..AlaPheThrGlnAlaAlaAspSerArgLeuGluMeta 476
      |||
3457 AAATACCCAGCACTTAACACTACAGGAGGATCAAGATTAACGCA 3506
476 spValGlyThr...ThrLeuGluProAlaAsp.....Thr 486
3507 CCAGTGTGATCCTTAACATCAATGCAAAAGATGCCAAATAGATGCTGCT 3556
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLy 503
      |||
3557 GCATCAGGTGACCGCAGTAGTAATGCAACTACGCAAGTGTCTG 3606
503 sLysAlaLysIleGluThrLysAlaIleThrSerLysAsnLeuThrLeuSerG 520
      |||
3607 TAAGGTG.....ACTGCGAAACCTCAAGCAGCTGAAATACACCG 3647
520 LysThrIleThrLeuLeuAspProThrGlyThrPheThrGlyAsn 534
3648 GGGATTTAAACACATTAATGCGTTAAATCATTTGCGAAAT 3691
seq_name: /cgn2_6/pcodata/2/lna/5D_COMB.seq:US-08-728-470-7
seq documentation block:
; Sequence 7, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SPOUNCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

```

; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ. ID NO.: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-728-470-7

alignment_scores:
  Quality: 271.50      Length: 648
  Ratio: 0.828        Gaps: 29
  Percent Similarity: 50.617  Percent Identity: 22.531

alignment_block:
  US-09-428-122-2 x US-08-728-470-7 ..
  Align seg 1/1 to: US-08-728-470-7 from: 1 to: 4287

23 ThrGluThrValLeuAspSerSerAlaSerPheAspGlyAsnLysAsnG 39
      |||
1955 ACAAAATTACCTTTAATGTAGCGGCTCATTTGACAAACATGCGGCTTC 2004
39 YAsnPheSerValArgLysSer..... 46
2005 AAACATTTCCATTGCCAGAGAGGGGCTAAATTTAAAGATATCAATACA 2054
47 .....GlnGluAspAlaGlyThrThrTyR.....Leu 55
2055 CCACTACCTTAATATATACCAACCACTGTGATACCACTTACCCACCAT 2104
56 PheLysGlyAsnValThrLeuGluAsnIlePheGlyThrGlyThrAlaI 72
      |||
2105 ATAAAGCAATATATCC.....AACAAATCAGGTGATTTGAATATATT 2148
72 eThrLysSerCys.....PheAsnAsn 80
      |||
2149 TGATTAATAAAAGCAGCGCTGAATTCGAATTTGGCGGCAATATCTCACAA 2198
80 hLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLeuPheGln 96
      |||
2199 AAGAAGCAATCTCACAAATTTCTTCTGATTAAGTAAATATATACCAATCAG 2248

```

```

97 ...ThrValAspIaGlyThrValaIaGlyAlaValaIaSerSerVa 112
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2249 ARAACAAATCAAAAGCAGCGCTTAAGAGGCGGCTTCTGATCAAGTGAAGC 2298
112 lValAspIysSerThrThr.....SerLeuSerA 118
   : : : : :
2299 AGAAAATGCTAACTAACTATTCACCAAGACTTAATAATTGGCAGAG 2348
119 .....PheIleGlyPheSerSerLeuSerPheIleAlaSerProGly 132
   : : : : : ||||| ||||| ||||| ||||| ||||| |||||
2349 ACCATAAATTTCAAGCTTTAATAAGCAGAAATTACAGCTAAATAATGCG 2398
133 SerSerIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerLe 149
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2399 AGTGAATTAATACTATGCG.....AATCTAGCGCTGGTAATGCG 2436
149 uSerLeuThrLys..... 153
   : : : : :
2437 TGATGCTAAAAAAGTGAAGCTTTTGACAAGCTTAAAGATTCAAAATCTCGA 2486
154 .....AsnValSerLeuLeuPheSerLysAsnPheSerThrAsp 166
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2487 CTGACGGCTCAAAATGTAACACTA.....AATGCGAAGTGAATAAGCTCT 2530
167 AsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMe 183
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2531 AATGCTAGTGAAGC...AATGCTGTAATGATTAACAGCAGCGGTTAAACAT 2577
183 tSerAla.....LeuPheSerGluAsnThrSerSerLysLys.... 195
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2578 TTCGCGAAAAGATGTAACGGTAAACAATAACGTACTACTCCCAAGACAA 2627
196 .....GlyGlyAlaIleGlnThrSerAspAlaLeuThr 206
2628 TAAATATCTCTGCCGACGACGAAGAATGTACAAACCAAGAGGACCAACT 2677
207 lIleThrGlyAsnGlnGlyIuValSerPheSerAspAsnThrSerSerAs 223
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2678 ATCAATGCAACCAAGCAGCGCTGGAAGTAAGTCTCAAAATGGTACAT 2727
223 pSerGlyAlaAlaIlePheThrGluAlaSerValThrIleSerAsnAsnA 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2728 TAAAGGCACAC.....ATTACCTCCGCAAAATGTAACTGACACAGACAG 2771
240 lAlaValSerPheIleAspAsnLysValThrGlyAlaSerSerSerThr 256
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2772 AAAATCTTGTTACACAGAGAAATGCTGCTAATGCAACACAGCGGACAA 2821
257 .....ThrGlyAspMetSerGlyGlyAlaIleCysAl 267
2822 GTAAACATTACTACAAAACAGGGGATATTAAAGTGA.....AT 2862
267 atYlYsThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetL 284
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2863 TGAATCAACTCTCCGTAATGTAATATATTACAGCAGCGGCAAT...ACAC 2909
284 euLeuPheSerAsnAsnThr..... 290
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2910 TTAAGGTAAGTAATATCACTGCTCAAGATGTAAACAGTAAACAGCGGATGA 2959
291 .....SerThrAlaIaGlyAlaIleTyValLys..... 301
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2960 GGAGCTTGACACTACAGCAGCGCTCAACCATTAAGTCGCAACAGGCAA 3009
302 .....LysLeuGluLeuA 306
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3010 TGCAAAATATTACCAACCAAGAGTGATATCAACGGTAAAGTTGAATGCA 3059
306 laseGlyGlyLeuThrIlePheSerThrAsnSerValAsnGlyLysThr 322
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3060 GCTCCGGCTCTGTAAACACTTGTTCACACTGGAGCAACTCTTCTGTAGAT 3109

```

```

323 AlaProLysGlyGlyAlaIleAlaIle...GluAspSerGlyLeu... 337
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3110 AATATTATGAGTAACACACTACTACTATTACTGCGATAGCGGTAAATTAC 3159
338 .....SerLeuSerA 341
3160 CTCACACTAGTCTCTCAATTAATGGACTAATAGTAAACCACTCA 3209
341 lAspSerGlyAspIle.....ValPheLeuGlyAsnThrValThrSer 355
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3210 GCCAATCGCGGATTAAGTAAGTAACTTCTGCTAATACAGTAATGTT 3259
356 ThrThrProGlyThrAsnArgSerSerIleAspLeuGlyThrSerAla 372
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3260 ACA.....GCAACACTGCTGATTACTATTGGAATAATAGTCANA 3300
372 SMetThrAlaLeuArgSerAlaIaGlyArgAlaIleTyPheTyAspP 389
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3301 AGTGAAGCGAAAATGGAAGCTCA..... 3325
389 rollerThrThrGlySerSerThrThrValThrAspValLeuLysValAsn 405
3325 ..... 3325
406 GluThrProAlaAspSerAlaLeuGlnTyThrGlyAsnIleIlePheTh 422
3325 ..... 3325
422 rGlyGlyLysLeuSerGluThrGluAlaAlaAspSerLysAsnLeuThr 439
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3326 .....ACCTTAACCTGCTGAATCAAGCAACCTTAACCA 3356
439 eryLysLeuGlnProValThrLeuSerGlyGly.....ThrLeuSer 453
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3357 CCCAAACAGCGCTGTGACTTACCTCAAGCAATGCTACAGCAACCTTACA 3406
454 LeuLysHis.....GlyValIThrLe 460
4407 GCCAAGGATAGCAGTATGCGAGAAACATTATCTCTAATGTGAGAGTT 3456
460 uGlnIleThrGln...AlaPheThrGlnAlaIleAspSerArgLeuGlnMetA 476
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3457 AATATCCACAGCGCACTTAACACTACAGGAGATTAACAGATTAAACCA 3506
476 sPValGlyThr...ThrLeuGluProAlaAsp.....Thr 486
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3507 CCAGTGTACTTAAACATCAATCAAAAGATGCCAATTAAGTGTGCT 3556
487 SerThrIleAsnAsnLeuValIleAsnIleSerSerIleAspGlyAla 503
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3557 GCATCAGGTGACCCACAGATGAATGCAACTAAGCAAGTGTGCTG 3606
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSerG 520
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3607 TAAAGTG.....ACTCGAAACCTCAAGCAGCGTGAATATCAACCG 3647
520 lYThrIleThrLeuLeuAspProThrGlyThrPheTyGluAsn 534
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3648 GGGATTTAACACAATAATAGGTTAATATATCATTTGCGAAAT 3691
seq_name: /cgn2_6/ptodata/2/lna/5D_COMB.seq:US-08-617-697-7
seq_documentation_block:
: Sequence 7, Application US/08617697
: Patent No. 5977336
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```



```

338 .....SerLeuSerA 341
3399 CTCACAGTAGTCTACATTAAGGACCTAATAGTGTACACCACTCAA 3988
341 laaSPserGlyaspIle.....ValPheLeuGlyAsnThrValThrSer 355
3989 GCCAATCGAGCGCATGTGAGGTACATTTCTGTAATACAGTAATGTT 4038
356 ThrThrProGlyThrAsnArgSerSerLeuAspLeuGlyThrSerAla 372
4039 ACA.....GCACGACTGGTGATTTAACTATGTGAATAGTGCAAA 4079
372 smethrAlaLeuArgSerAlaAlaGlyArgAlaIleTyrPheTyrAsp 389
4080 AGTTGAAAGCAAAATGAGCTGCA..... 4104
389 rolIethrThrGlySerSerThrThrValThrAspValLeuLysValAsn 405
4104 ..... 4104
406 GluThrProAlaAspSerAlaLeuGlnTyrThrGlyAsnIleIlePhe 422
4104 ..... 4104
422 rGlyGlyLeuSerGluThrGluAlaAlaAspSerLysAsnLeuThrS 439
4105 .....ACCTTAAGTCTGATCAGCAATTAACA 4135
439 erLysLeuGlnProValThrLeuSerGlyGly.....ThrLeuSer 453
4136 CCCAACAGCGCTAGCATTTACTCTCAAGCATGTGACAACTCTTACA 4185
454 LeuLysHis.....GlyValThrIle 460
4186 GCCAAGGATAGCATTCGACGAAACATTAAATGCTAATGTCAGCT 4235
460 uGlnThrGln.....AlaPheThrGlnGlnAlaAspSerArgLeuGluMet 476
4236 AATATACCAAGCAGCATTAACTACTACAGGAGATTCAAAGATTAAAGCAA 4285
476 spValGlyThr...ThrLeuGluProAlaAsp.....Thr 486
4286 CCAAGTGGTCTTAACAAATCAATCAAAAGATGCCAAATAGATGCT 4335
487 SerThrIleAsnLeuValIleAsnIleSerSerIleAspGlyAla 503
4336 GCATCAGGAGCAGCAGCAGTAACTCAACTACGCAAGTGGCTCTCG 4385
503 sLysAlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeuSer 520
4386 TAACTG.....ACTGCAAAACCTCAAGCAGGTGAATATCACCG 4426
520 lYThrIleThrLeuLeuAspProThrGlyThrPheTyrGlyAsn 544
4427 GGAATTAACACAAATAATGCTTAATATCATTTGCAAAAT 4470
seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:us-08-038-682-8
seq_documentation_block:
; Sequence 8, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltre, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia

```

```

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4702 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-038-682-8

```

```

alignment_scores:
Quality: 270.00 Length: 644
Percent Similarity: 49.845 Percent Identity: 22.671

```

```

alignment_block:
US-09-428-122-2 x US-08-038-682-8

```

```

Align seg 1/1 to: US-08-038-682-8 from: 1 to: 4702

```

```

27 LeuAspSerSerAlaSerPheAspGlyAsnLysAsnGlyAsnPheSerA 43
2537 CTAAACATCACCGGACCTTACCAACAAACGTCACCGCAACATTATAT 2586
43 lArgLysSerGlnGluAspAlaGlyThrThrTyrLeuPheLysGly... 58
2587 AAAACAAGAGGCTGAACACTCCAAAGGATATATCAATAAGGCTGT 2636
59 ..AsnValThrLeuGluAsnIleProGlyThrGlyThrAlaIleThrLys 74
2637 TAAATATCACT...ACTTAACGCTCAAGGCACTCAAAAACCATTTATAC 2683
75 SerCysPheAsnAsnThrLysGlyAspLeuThrPheThr..... 87
2684 GGAATATATACTTAACGAAGAAAGGCACTTAACATCAAGAAATTTAAAC 2733
88 .....GlyAsn.....GlyAsn 92
2734 CGAGCGCAAAATCCAATTCGCGCAATATCTCAGCAAAAGAGCAATC 2783
92 erLeuLeuPheGln.....ThrValAsp 99
2784 TCACAAATTTCTCTGATAAAGTAATATTACCAATCAGATACATCAA 2833
100 AlaGlyThrValAlaGlyAlaValaValaAsnSerValValaLysLys 116
2834 GCAAGCCTTAAGGGGGCGGTCTGATTCAGTGAAGGCAAGAAATGCTTA 2883
116 rThrThr.....PheI 120
2884 CTTAACTATTCAAAACCAAGATTAAATTCGACAGAGACCTAAATATT 2933
120 leGlyPheSerSerLeuSerPheIleAlaSerProGlySerSerIleThr 136
2934 CAGGCTTAATAAAGCAAGAAATTACAGCTAAATAATGGCAGTATTAACT 2983

```



```

137 ThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSerLeuThrLy 153
154 snValSerLeuLeuPheSerLysAsnPheSerThrAspAsnGlyAla 170
3072 ATGTAAACACTA.....AATAGCCAGTGAAGAAAGCTAATAGTAGAGC 3115
171 IleThrAlaLysThrLeuSerLeuThrGlyThrThrMetSerAla.... 185
3116 ..AATGCTGTGAATGATTAACGACCGGCTTAACCATTCCTCCGAAAGA 3162
186 ...LeuPheSerGluAsnThrSerSerLysLys.....A 195
3163 TGTAACGGTAACAATACGTTACCTCCACAGACAAATAAATCTCTG 3212
196 ....GlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAsn 210
3213 CCGCAGCAGGAATGTAAACAACCAAGAGGACACATCATCATGCAACC 3262
211 GlnGlyLysValSerPheSerAspAsnThrSerSerAspSerGlyAlaAl 227
3263 ACAGGCGCGGTGAAGTAACTGCTCAAAATGCTACAAATTAAAGCAAC. 3310
227 AliePheThrGluLysValThrIleSerAsnAlaLysValSerP 244
3311 ...ATTACCTCGCAAAATGTAAACAGTACAGCAGCAAGAAATCTGTGA 3356
244 heIleAspAsnLysValThrGlyAlaSerSerThr..... 256
3357 CCACAGGAGATGCTGCTAATATCAACACGCGACAGTAAACATTAGT 3406
257 .....ThrGlyAspMetSerGlyGlyAlaIleGlySalaryLysThrse 271
3407 ACAAAAACAGGGGATTTAAAGGTGA.....ATTGATTAACCTTC 3447
271 rThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSera 288
3448 CGGTAAATGTAAATATACAGCGAGCGCAAT..ACAATTAGGTAAATGA 3494
288 snAsnThr.....Ser 291
3495 ATATCACTGTGCAAGATGTAACTAAACAGCGATGACAGACCTTGACA 3544
292 ThrThrAlaGlyGlyAlaIleTyrValLys..... 301
3545 ACTACACAGCGCTAACCATTTAGTCGCACAAAGCAGCAATGCAATATTAC 3544
302 .....LysLeuGluLeuAlaSerGlyL 310
3595 AACCAAAACAGGTGATATCAACGGTAAAGTTGATTCAGCTCCGCTGTG 3644
310 eutThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaProLysGly 326
3645 TAAACACTGTGTCAACTGAGCACTCTTGCTGTAGTAATATTTCAGGT 3694
327 GlyAlaIleAlaIle...GluAspSerGlyGluLeu..... 337
3695 AACACGTCTTACTTACTCGCGATAGCGGTAAATTAACCTCCACAGTAGG 3744
338 .....SerLeuSerAlaAspSerGly 345
3745 TTCTACATTAATAGGACTAATAGTGTAAACACCTCAAGCAATCAGCG 3744
345 spIle.....ValPheLeuGlyAsnThrValThrSerThrThrProGly 359
3795 ATATTGAAGGTACAAATTTCTGTGTAATACAGTAATGTTACA..... 3835

```

```

360 ThrAsnArgSerSerIleAspLeuGlyThrSerAlaLysMetThrAlaIe 376
3836 GCACAGCACTGTGATTTACTATTGAAATAGTCAAAAGTTGAAGCGAA 3885
376 uArgSerAlaAlaGlyArgAlaIleTyrPheTyrAspProIleThrThg 393
3886 AAATGAGCTGCA..... 3898
393 LysSerSerThrThrValThrAspValLeuLysValAsnGluThrProAla 409
3898 ..... 3898
410 AspSerAlaLeuGlnTyrThrGlyAsnIleIlePheThrGlyLysLe 426
3898 ..... 3898
426 uSerGluThrGluAlaAlaAspSerLysAsnLeuThrSerLysLeuLeu 443
3899 .....ACCTTAACGTGCTGAATCAGCAATTAACCCCAACAGGCT 3941
443 InProValThrLeuSerGlyGly.....ThrLeuSerLeuLysHis... 456
3942 CTAGCATTTACCTCAACGAATGTACAGCAACCTCTTACAGCAGCATAGC 3991
457 .....GlyValThrLeuGlnThrGln... 463
3992 AGTATGCGACGAAACATTAATGCTGCTAATGTGACGTTAAATCCACAG 4041
464 .AlaPheThrGlnGlnAlaAspSerArgLeuGlnMetAspValGlyThr 479
4042 CACTTTAATCACTACAGGCGATTCAAAGATTACGCAACACAGGTGACCT 4091
480 ..ThrLeuGluProAlaAsp.....ThrSerThrIleAsn 490
4092 TAAACAATCAATGCCAAAGATGCCAAATTGATGCTGCTCATCAGCTGAC 4141
491 AsnLeuValIleAsnLysSerSerIleAspGlyAlaLysLysAlaLysI 507
4142 CGCACAGTAAATATGCACTTAACGCAAGTGGCTCTGCTGTAACGTG.... 4186
507 eGluThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThrL 524
4187 ....ACTGCCAAACCTCAACGACGCTGAATATACACCGGAGATTAAACA 4232
524 euleuAspProThrGlyThrPheTyrGluAsn 534
4233 CAATTAATGGTTAAATATCATTTTCGGAAT 4264
seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-302-832-8
seq_documentation_block:
; Sequence 8, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; SUITE: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832

```

```

: FILING DATE: 16-SEP-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US Pct/us93/02166
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Belkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-404
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0810
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4702 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
:
US-08-302-832-8

alignment_scores:
      Quality: 270.00      Length: 644
      Ratio: 0.841      Gaps: 29
      Percent Similarity: 49.845      Percent Identity: 22.671

alignment_block:
US-09-428-122-2 x US-08-302-832-8 ...

Align seg 1/1 to: US-08-302-832-8 from: 1 to: 4702

27 LeuAspSerSerAlaSerPheAspGlyAsnLysAsnPheserVa 43
|||||:.....:|||||:.....:|||||:.....:
2537 CTAACATCACCAGCAGCTTACCAACACGTCACCGCACTTAAAT 2586
43 IARGGUSERGNGLUASPLAAGLYTHRTHRLEUPHELYSGLY.... 58
:.....:|||||:.....:
2587 AAAACAAGAGTGTTAAACCTCCAGCGGAGATATCAATTAAGTGCT 2636
59 . AsnValThrLeuGluAsnLeuProGlyThrAlaIleThrLys 74
|||||:.....:|||||:.....:
2637 TAAATATCACT...ACTAAGCCTCAGCAGCTCAAAAACCATTTATTAAC 2683
75 SerCysPheAsnAsnThrLysGlyAspLeuThrPheThr..... 87
:.....:|||||:.....:
2684 GGAATATTAAGTAAAGGCGACTTAAACATCAAGATATTTAAAGC 2733
88 .....GlyAsn.....GlyAsn.....GlyAsn 92
|||||:.....:|||||:.....:
2734 CGAGCGCGAATCCAAATTTGGCGCAATATCTCAGAAAAGAGCGCATC 2783
92 erLeuLeuPheGln.....ThrValAsp 99
:.....:|||||:.....:
2784 TCACAAATTTCTCTGATAAGTAATATTAACATCAGATAACAATCAA 2833
100 AlAGLYThValAlAGLYAlaAlaValAsnSerSerValValAspLys 116
|||||:.....:|||||:.....:
2834 GCAGGGGTTGAAGGGGGGCTTGATTCAGTGAGGCGAAMAATGCTAA 2883
116 rThrThr.....PheI 120
:.....:|||||:.....:
2884 CTTAACTATTCAACCAAGAGTTAAATTTGGCGAGGAGACTAAATTT 2933
120 IeGLYTheserSerLeuSerPheIleAsnProGlySerSerIleThr 136
|||||:.....:|||||:.....:
2934 CAGGCTTTATTAAGCAGAAATTAACGCTTAATAATGGCAGTGAATTAAC 2983
137 ThGlyLysGlyAlaValSerCysSerThrGlySerLeuSerLeuThrly 153
|||||:.....:|||||:.....:

```

```

2984 ATTGCG.....AATGCTAGCGGTGTAATGCTATGCTAATAAA 3021
153 S.....A 154
3022 AGTACCTTTGCACACGTTAAAGATTCAAAAATCTGCAGTCAGGCACAA 3071
154 snValSerLeuLeuPheSerLysAsnPheserThrAspAsnGlyGlyAla 170
|||||:.....:|||||:.....:
3072 ATTTAACACTA.....AATAGCAGTGAAGAAACGCTTAATGTAAGTAC 3115
171 IleThrAlaLysThrLeuSerLeuThrGlyThrThrMetSerAla..... 185
:.....:|||||:.....:
3116 ..AATGCTGTAATGATTAACAGCAGCCGTTTAACCATTTCCGCAAAAGA 3162
186 ....LeuPheSerGluAsnThrSerSerLysLys..... 195
:.....:|||||:.....:
3163 TGTACGCTTAACAAATACGTACTCTCCACACACAAATATATCTGTG 3212
196 ....GlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAsn 210
|||||:.....:|||||:.....:
3213 CCGCAGCAGGAAATGTACACACCAAGAGGACACACTAATCAATGCAACC 3262
211 GlnGlyLysValSerPheSerAspAsnThrSerSerAspSerGlyAlaAl 227
|||||:.....:|||||:.....:
3263 ACAGCAGCGCTGCAAGTACTCTCAAAATGTACAAATTTAAAGGCACAC.. 3310
227 alIePheThrGluAlaSerValThrIleSerAsnAsnAlaLysValSerP 244
:.....:|||||:.....:
3311 ...ATTACCTCCCAAAATGTACAGAGTGACAGCAACAGAAATCTGTGA 3356
244 helIeAspAsnLysValThrGlyAlaSerSerThr..... 256
|||||:.....:|||||:.....:
3357 CCACAGAGAAATGCTGTCATTATTAACACACGAGCAGACAGTAACATTA 3406
257 ....ThrGlyAspMetSerGlyValAlaIleCysAlaThrLysThrSe 271
|||||:.....:|||||:.....:
3407 ACAAAAACAGGAGATTTAAAGTGA.....ATTGAATCAACTTC 3447
271 rThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerA 288
|||||:.....:|||||:.....:
3448 CGGTAATGTAATATTAACAGCAGCGGCAT...ACACTTAAGTAGTA 3494
288 snAsnThr.....Ser 291
|||||:.....:|||||:.....:
3495 ATATCACTGTCAGATGTACAGTACAGCGGATGCGAGCGCTTGACA 3544
292 ThrThrAlaGlyAlaIleThrValLys..... 301
|||||:.....:|||||:.....:
3545 ACTACACAGCGCTCAACCATTTAGTCGACACACAGCAATGCCAATATTAC 3594
302 .....LysLeuGluLeuAlaSerGlyLys 310
|||||:.....:|||||:.....:
3595 AACCAAAACAGTGATATCAACGTAAGTGAATCCAGCTCCGGCTCG 3644
310 euThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaProLysGly 326
|||||:.....:|||||:.....:
3645 TTAACCTTTGCACTGAGCACTCTGCTGTAGGTAATATTAGGT 3694
327 GlyAlaIleAlaIle...GluAspSerGlyLeu..... 337
:.....:|||||:.....:
3695 AACACTGTTACTATTACTGCGGATGCGGTAAATTAACCTCCACAGTAG 3744
338 .....SerLeuSerAlaAspSerGlyVa 345
:.....:|||||:.....:
3745 TTTCACATTAATGAGACTAATAGTGAACACCTCCAGCAATCAGCGG 3794
345 spIle.....ValPheLeuGlyAsnThrValThrSerThrProGly 359
|||||:.....:|||||:.....:
3795 ATATGAGGTACAAATTTCTGGTAATACAGTAATGTTAA..... 3835
360 ThrAsnArgSerSerIleAspLeuGlyThrSerAlaLysMetThrAla 376
|||||:.....:|||||:.....:
3836 GCAAGCACTGTGATTACTTACATTGGAATAGTGCACAAAGTGAACGCA 3885

```

```

376 uargSerAlaIagIyArghAlaIleTyrrhPheTyrrAspProIlethrThrg 393
      ::::::::::::::
3886 AATGGAGCTGCA..... 3898
393 lySerSerThrThrValThrAspValLeuLysValAsnGluThrProAla 409
3898 ..... 3898
410 AspSerAlaLeuGlnTyrrThrGlyAsnIleIlePheThrGlyLysLe 426
3898 ..... 3898
426 uSerGluThrGluAlaAlaAspSerLysAsnLeuThrSerLysLeuL 443
      ||| ::::::::::::::
3899 .....ACCTTAAGCTGCTGATATGAGCAATTTACCAACCAAGAGCT 3941
443 InProValThrLeuSerGlyGly.....ThrLeuSerLysHis... 456
      ::::::::::::::
3942 CTACATTACTCTCAAGCAATGTCAGACAACTCTTACAGCAAGATAGC 3991
457 .....GlyValThrLeuGlnThrGln.. 463
3992 AGTATCCAGAGAAACATTAAATGCTGTATATGCTTAATATCCACAG 4041
464 .AlaPheThrGlnGlnAlaAspSerArgLeuGluMetAspValGlyThr. 479
      ::::::::::::::
4042 CACTTACTACTACTACAGGGATTAAGATTAAACGACCAACAGTGTACT 4091
480 .ThrLeuGluProAlaAsp.....ThrSerThrIleAsn 490
      ||| ::::::::::::::
4092 TAACAATCAATGCMAAAGATGCCAAATTAGATGCTGCATCAGTGAGC 4141
491 AsnLeuValIleAsnIleSerSerIleAspGlyAlaLysAlaLysI 507
      ::::::::::::::
4142 CGCACACTAGTAATGCACTACACGACAGTGGCTCTGTACAGTG..... 4186
507 egluThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThrL 524
      ||| ::::::::::::::
4187 ....ACTGCAAAACCTCAAGCAGCGATATCACCAGCGGATTTAACA 4232
524 euLeuAspProThrGlyThrPheTyrrGluAsn 534
      ::::::::::::::
4233 CAATAAATGGGTAAATATCATTCGGAATAAT 4264

seq_name: /cgn2_6/ftodata/2/lna/5C_COMB.seq:us-08-530-198-8

seq_documentation_block:
; Sequence 8, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-530-198-8

alignment_scores:
      Quality: 270.00      Length: 644
      Ratio: 0.841      Gaps: 29
      Percent Similarity: 49.845      Percent Identity: 22.671

alignment_block:
US-09-428-122-2 x US-08-530-198-8 ..

Align seg 1/1 to: US-08-530-198-8 from: 1 to: 4702

27 LeuAspSerSerAlaSerPheAspGlyAsnLysAsnGlyAsnPheserVa 43
      ||||| ::::::::::::::
2537 CTAAACATCAGCGGACCTTTCCAAACAAGCGTACCGCCACATTAATAT 2586
43 lArgGluSerGlnGlnAlaSpAlaGlyThrThrTyrrLeuPheLysGly... 58
      ::::::::::::::
2587 AAACACAGAGTGTAAACCTCAAGCGCATATTATCAATAAAGTGTGTT 2636
59 .AsnValThrLeuGluAsnIleProGlyThrGlyThrAlaIleThrLys 74
      ||||| ::::::::::::::
2637 TAAATATCACT...ACTAACGCTCAGGCACATCAAAAACCATTTATAC 2683
75 serCyPheAsnAspThrLysGlyAspLeuThrPheThr..... 87
      ::::::::::::::
2684 GGAATATATCACTACGAAAGAGCGCATTAACATCAACATATTAAAGC 2733
88 .....GlyAsn.....GlyAsn 92
      |||||
2734 CGACGCCGAATCCAAATGCGGCAATATCTCACAAAAGAGCAATC 2783
92 erLeuLeuPheGln.....ThrValAsp 99
      ::::::::::::::
2784 TCACAATTCTCTCTGATTAAGTAATATTACCAATCAGATACAAATCAA 2833
100 AlaGlyThrValAlaGlyAlaAlaValaLysSerSerValValaLysSe 116
      ||||| ::::::::::::::
2834 GCGGGGTTGAAGGGGGGCTTCTGATTCAAGTGAAGGCAAAAATGCTAA 2883
116 rThrThr.....PheI 120
      ::::::::::::::
2884 CCAATCTATTCAACCAAGAGTTAAATTGCGAGAGACCTAATATT 2933
120 leglyPheSerSerLeuSerPheIleAlaSerProGlySerSerIleThr 136
      ||||| ::::::::::::::
2934 CAGGCTTTAATTAAGCAGAAATTAACAGCTAAATAATGCGCATATTACT 2983
137 ThrGlyLysGlyAlaValSerCySerThrGlySerLeuSerLeuThrLys 153
      ||| ::::::::::::::
2984 ATTGGC.....AATGCTACGGGTGTAATGCTGATGCTAAAAA 3021
153 s.....A 154
      |
3022 AGTACATTTGACAAGGTTAAGATTCAAAAATCTGACTGACGGTCA 3071
154 snValSerLeuLeuPheSerLysAsnPheserThrAspAsnLysLysAla 170
      ||||| ::::::::::::::

```

```

3072 ATGTACACTA.....AATACGACGAGTGAACCTAATGCTAGTAC 3115
171 IleThrAlaLysThrLeuSerLeuThrGlyThrThreSerAla..... 185
3116 ...AATGCTGTGAATATGATACAGACGCGTTTAACCATTTCCCAAAAGA 3162
186 ....LeuPheSerGluAsnThrSerSerLysLys..... 195
3163 TGTACGCGTAACATATACCTTACTCTCCACAGACATATAATATCTCTG 3212
196 ....GlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAsn 210
3213 CCGCAGAGAGAAATGTACACACCAAGAGCAGACACTATCAATGCAACC 3262
211 GlnGlyGluValSerPheSerAspAsnThrSerSerAspSerGlyAlaAl 227
3263 ACAGGCGAGCTGTGAGTACCTGCTCAAAATGTACAAATTAAGGCAAC.. 3310
227 aIlePheThrGluAlaSerValThrIleSerAsnAlaValSerP 244
3311 ....ATTACCTCGCAAAATGTACAGTACAGCAGACAGAAATCTTGTA 3356
244 helLeuAspAsnLysValThrGlyAlaSerSerThr..... 256
3357 CCACAGAGATGCTGTCTATTATTCACACGCGCAGACAGTAAACATTAGT 3406
257 .....ThrGlyAspMetSerGlyGlyAlaIleCysAlaTyrIysThre 271
3407 ACAAAAACAGCGGATATTAAGGTGA.....ATTGAATCAACTTC 3447
271 rThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuPheSerA 288
3448 CGGTAATGTAAATATATACAGCGCGCAAT..ACACTTAAGTAAAGA 3494
288 spAsnThr.....Ser 291
3495 ATATCACTGTCTAGATGTACAGTAAACAGCGGATGCAGAGGCTTGACA 3544
292 ThrThrAlaGlyAlaIleTyrValLys..... 301
3545 ACTACAGCGGCTCAACCATTTAGTCGACACAGCAAGCAATTAATTAC 3594
302 .....LysLeuGluLeuAlaSerGlyL 310
3595 AACCAAAACAGTGATATCAACGTAAGTGAATCCAGCTCCGGCTCG 3644
310 euThrLeuPheSerArgAsnSerValAsnGlyThrAlaProLysGly 326
3645 TAACTCTTGTGCACTGAGCAGCACTCTGCTGAGTAATATTTCAGGT 3694
327 GlyAlaIleAlaIle...GluAspSerGlyGluLeu..... 337
3695 AACACTGTACTATTAATCGGATAGCGGTAAATTAACCTCCACAGTAGG 3744
338 .....SerLeuSerAlaAspSerGlyA 345
3745 TTCTACATTAATGGCACTAATAGTGTACACCACTCAAGCATCAGGG 3794
345 spLle.....ValPheLeuGlyAsnThrValThrSerThrProGly 359
3795 AATTTGAGTACATTTCTGTATACAGTAATGTACA..... 3835
360 ThrAsnArgSerSerLeuAspLeuGlyThrSerAlaLysMetThrAlaI 376
3836 GCAAGCACTGCGTATTTAACTATTGGAATGTGCACAAAGTTAAACGCG 3885
376 uArgSerAlaAlaGlyArgAlaIleTyrPheTyrAspProIleThrThc 393
3886 AATGAGCGTGA..... 3898
393 LysSerSerThrThrValThrAspValLeuLysValAsnGluThrProAla 409
3898 ..... 3898

```

```

410 AspSerAlaLeuGlnTyrThrGlyAsnIleIlePheThrGlyGluLysLe 426
3898 ..... 3898
426 uSerGluThrGluAlaAlaAspSerLysAsnLeuThrSerLysLeuG 443
3899 .....ACCTTAAGTGTGATCAGGCAATTAACCAACCAACGAGGT 3941
443 InProValThrLeuSerGlyGly.....ThrLeuSerLeuLysHis... 456
3942 CTAGCATTAAGTCAAGCAATGTCTCAGACACTCTTACAGCGCAAGGTAGC 3991
457 .....GlyValThrLeuGlnThrIn.. 463
3992 AGTATCGCAGAAACATTAATGCTGCTAATGTGCGTTAATCCACAG 4041
464 .AlaPheThrGlnAlaAspSerArgLeuGluMetAspValGlyThr. 479
4042 CACTTACTACTACAGGGGATTCAGGATTAACGCAACAGGTGACT 4091
480 ....ThrLeuGluProAlaAsp.....ThrSerThrIleAsn 490
4092 TAACATCAATGCCAAAGATGCCAAATTAGATGCTGCTGATAGGTGAC 4141
491 AsnLeuValIleAsnIleSerSerIleAspGlyAlaLysAlaLysTl 507
4142 CGCACAGTAAATGCACTACACCAAGTGGCTGTGTAAGTGA..... 4186
507 eGluThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThrL 524
4187 ....ACTGCAAAACCTCAAGCAGCGTGAATATCACCGGAGATTAAACA 4232
524 euLeuAspProThrGlyThrPheTyrGluAsn 534
4233 CAATAATGGTTAATATATCATTTCCGCAAAAT 4264

seq.name: /cgn2_6/prodata/2/ina/5C_COMB.seq:US-08-469-880-8

seq.documentatation_block:
: Sequence 8, Application US/08469880
: Patent No. 5876733
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J.
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,880
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA: US 08/302,832
: APPLICATION NUMBER:
: FILING DATE: 16-SEP-1994

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Berckstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4702 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-8

alignment_scores:
Quality: 270.00      Length: 644
Ratio: 0.841        Gaps: 29
Percent Similarity: 49.845      Percent Identity: 22.671

alignment_block:
US-09-428-122-2 x US-08-469-880-8 ..

Align seg 1/1 to: US-08-469-880-8 from: 1 to: 4702

27 LeuaspSerSerAlaSerPheaspGlyAsnLysasnGlyaspPheSerVa 43
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2537 CTAACATCTACCGGACCTTACCACACAGCATCCGCCAACATTATAT 2586
43 LavgGlueSerGlnLuspaLagIThrThrTyLeuPheLysGly... 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2587 AAAACACAGGGGTGTAATACTCCAGGCGCATATTATCAATPAAGGTGT 2636
59 .AsnValThrLeuGluAsnIleProGlyThrGlyThrAlaIleThrLys 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2637 TAAATATCACT...ACTACGCGCTCAGGCACTCAAAAACCATTTATAC 2683
75 SerCysPheAsnAsnThrLysGlyAspLeuThrPheThr... 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2684 GGAATATTAATACTAAGAAAAGCGACTTAACATCAAGATATTAAAGC 2733
88 .....GlyAsn.....GlyAsnS 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2734 CGACGCCGAATCCAAATTCGCGCATATCTACAAAAGAAAGGCAATC 2783
92 erLeuLeupheGln.....ThrValasp 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2784 TCACAAATTTCTCTGATAAAGTAATATTACCAATCAGATACATCAAA 2833
100 AlagIyThrValAlaGlyAlaAlaValAsnSerSerValAlasplysSe 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2834 GCAGGCGTTAAGGGGGCGCTTCGATTCAAGTGAGGAGAAAATGCTAA 2883
116 rThrThr.....PheI 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2884 CTAACATATTCACAAACCAAGATTAAATTCGACGACGACCTAAATFTT 2933
120 leGlyPheSerSerLeuSerPheIleAlaSerProGlySerSerIleThr 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2934 CAGGCTTTAATAAGCAAGAAATTACAGCTAAATAATGACAGTGAATTACT 2983
137 ThrGlyIyGlyAlaValSerCysSerThrGlySerLeuSerLeuThrIly 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2984 ATTGGC.....AATGCTAGCGGTGTGTAATGCTGATGCTAAAAA 3021
153 S.....A 154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3022 AGTGACTTTTGACAGAGTTAAAGATTCAAAATCTGACTGACGCTGACA 3071
154 snValSerLeuLeupheSerLysAsnPheSerThrAspAsnGlyGlyAla 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

3072 ATGTACACTA.....AATAGCGAAGTGAAGAAAGCTTAATGCTAGTACG 3115
171 IleThrAlaLysThrLeuSerLeuThrGlyThrThrMetSerAla..... 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3116 ...AATGCTGTGTAATGATAACAGACCGGTTTAACCATTTCCGCAGAAAGA 3162
186 ...LeupheSerGluAsnThrSerSerLys..... 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3163 TGTAAAGGTAAACAAATACGTACCTCCCAACAGCAATTAATATCTGTG 3212
196 ....GlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAsn 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3213 CCGCAGCAGAAATGTAAACACCAAGAAAGGCAACATCTCAATGCAACG 3262
211 GlnGlyGluValSerPheSerAspAsnThrSerSerAspSerGlyAlaAl 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3263 ACAGCAGCGGTGGAAGTAACTGCTCAAAATGTACAAATTAAGGCANAC.. 3310
227 aIlePheThrGluAlaSerValThrIleSerAsnAlaLysValSerP 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3311 ...ATTACCTCGCAAAATGTAAACAGTGAACGACACAGCAAGAAATCTTGT 3356
244 heIleAspAsnLysValThrGlyAlaSerSerThr..... 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3357 CCACAGAGAAATGCTGTCAATTAATGCAACCGCGGCACTAAACATTACT 3406
257 .....ThrGlyAspMetSerGlyGlyAlaIleCysAlaTyLysThrSe 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3407 ACAAAACAGGGGTATTTAAGGTGA.....ATTGAATCAACTTC 3447
271 ThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuLeupheSerA 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3448 CGGTAAATGTAATATTACAGCGAGCGCAAT...ACACTTAAGGTAGTA 3494
288 snAsnThr.....Ser 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3495 ATATCACTGCTCAAGATGTACAGTAACAGCGGATGACGAGCTTGACA 3544
292 ThrThrAlaGlyAlaIleTyValLys..... 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3545 ACTACAGCAGGCTCAACATTTAGTGGACACAGCAATGCAAAATTTAC 3594
302 .....LysLeuGluLeuAlaSerGlyL 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3595 AACCAAAACAGTGATATCAAGGTAAAGTTGAATCAGCTCGGCTCG 3644
310 eutThrLeupheSerArGAsnSerValAsnGlyGlyThrAlaProLysGly 326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3645 TAACACTTGTTCGAACTGGAGCAACCTTGTCTAGTAATATTTCAGGT 3694
327 GlyAlaIleAlaIle...GluaspSerGlyGluLeu..... 337
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3695 AACACTGTACTATTCTCGCGGTAGCGTAATTTACCTCCACAGTAG 3744
338 .....SerLeuSerAlaaspSerGlyA 345
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3745 TCTACAAATTAATGGACTAATAGTGAACACCTCAACCAATCAGCGCG 3794
345 spIle.....ValPheLeuGlyAsnThrValThrSerThrProGly 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3795 ATATTAAAGCTACAAATTTCTGGTAATACAGTAATCTTACA..... 3835
360 ThrAsnArGSerSerIleaspLeuGlyThrSerAlaLysMetThrAlaLe 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3836 GCAGAGACAGTGATTTACTATTGGAATAGCGCAAAAGTTGAAGCGCA 3885
376 uArgSerAlaIleGlyArgAlaIleTyPheTyraspProIleThrThG 393
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3886 AAATGAGCTGCA..... 3898
393 LysSerSerThrThrValThrAspValLeuLysValAsnGlnThrProAla 409
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3898 ..... 3898

```

```
410 AsperalaleuGlnThrGlyAsnIleIlePheThrGlyIuysIe 426
3898 .....
426 userGluThrGluAlaAlaAspSerIysAsnLeuThrSerIysLeuG 443
3899 .....ACCTTAAGCTGATGATCAAGCAATTAACCAACCAAGGCT 3941
443 InProvalThrLeuSerGlyGly.....ThrLeuSerLeuIysGHS... 456
3942 CTACAGATTACCTCAAGCAATGTCAGACAACTTACGCCAAGATAGC 3991
457 .....GlyValThrLeuGlnThrGln... 463
3992 AGTATCCAGAAACATTAATGCTGCTAATGATGACGTTAAATACACAG 4041
464 .AlaPheThrGlnGlnAlaAspSerIysLeuThrSerIysGlyThr 479
4042 CACTTAAGCTGATGATCAAGGATTAAGCAATTAACCAAGGATGATCCT 4091
480 ..ThrLeuGluProIaAsp.....ThrSerThrIleAsn 490
4092 TAAACAATCAATGCAAAAGATGCCAATTAATGATGCTGATCAAGGTGAC 4141
491 AsnLeuValIleAsnIleSerSerIleAspGlyAlaIysIysAlaIysI 507
4142 CGCAGCTAGTAAATGCAATCAACCAAGTGGCTGCTGTAACGTG.... 4186
507 egluThrGlyAlaThrSerIysAsnLeuThrLeuSerGlyThrIleThr 524
4187 ...ACTGCAAAACCTCAAGCAGCGGTGATATCACCGGGGATTTAAACA 4232
524 euleuAspProthrGlyThrPheIyrgIuAsn 534
4233 CAATAATGGTAAATATCATTTGCGAATAAT 4264

seq.name: /cgn2_6/ptodata/2/ina/50_COMB.seq:US-08-728-470-8

seq.documentation_block:
; Sequence 8, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
```

```
ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-728-470-8

alignment_scores:
Quality: 270.00 Length: 644
Ratio: 0.841 Gaps: 29
Percent Similarity: 49.845 Percent Identity: 22.671

alignment_block:
US-09-428-122-2 x US-08-728-470-8 ..

Align seq 1/1 to: US-08-728-470-8 from: 1 to: 4702

27 leuAspSerSerAlaSerPheAspGlyAsnIysAsnGlyAsnPheSerIva 43
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2537 CTAAACATCAACCGGCGCTTATCCACACACGATACCGCCACATTAATAT 2586
43 IargGluSerGlnGluAspAlaGlyThrThrIyrluPheIysGly.... 58
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2537 AAAACAAGAGGTGTTAAATCCCAAGCGATATTATCAATATAAGGTGCTT 2636
59 ..AsnValThrLeuGluAsnIleProGlyThrGlyThrAlaIleThrIys 74
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2637 TAAATATCACT...ACTAACCGCTCAGCAGCATCAAAAACCATTTATAC 2683
75 SerCysPheAsnAsnThrIysGlyAspLeuThrPheThr..... 87
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2684 GGAATATATCAATCAAGAAAGGCGACTTAAACATCAAGATATTATAAGC 2733
88 .....GlyAsn.....GlyAsn 92
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2734 CGAGCGCGAAATCCAAATGGCGGATATCTCAACAAAGAGGCGATC 2783
92 erLeuLeuPheGln.....ThrValAsp 99
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2784 TCACAAATTTCTTGATTAAGTAATATTACCAATCAATCAATCAATCAAA 2833
100 AlaGlyThrValAlaGlyAlaValAsnSerSerValValAspIysSe 116
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2834 GCAGCGGTGGAAGGGGGGCTTCTGATTCAAGTGAAGGCAAGAAATGCTAA 2883
116 rThrThr.....PheI 120
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2884 CTAACTATTCAACCAAGATTAATGTCAGAGAGACCTAAATATT 2933
120 leuGlyPheSerSerLeuSerPheIleAlaSerProGlySerSerIleThr 136
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2934 CAGGCTTATTAAGCAAGAAATTAACAGCTAAATAGCAGATATTAACT 2983
137 ThrGlyIysGlyAlaValSerCysSerThrGlySerLeuSerLeuThrIly 153
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
2984 ATTGGC.....AATGCTAGCGGTGTAATGCTGATGCTAAATAA 3021
153 S.....A 154
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
3022 AGTGACTTTGCAAGGTTAAAGATTCAAAATCTGCACTGAGCGGTCA 3071
154 snValSerLeuLeuPheSerIysAsnPheSerThrAspAsnIyGlyAla 170
||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
```

3072 ATGTACACTA.....ATAAGCAAGTCAAAACGCTAATGCTAGTAGC 3115
 171 lIethrAlaLysThrLeuSerLeuThrGlyThrMetSerAla.... 185
 3116 ..AATGCTGAATGATACAGCACCGGTTTAACCATTTCCGCAAGA 3162
 186 ...LeuSerGluAsnThrSerLysLys..... 195
 3103 TGTACGGTAACATTAACGTTACTCCACAGACAAATAATATCTCTG 3212
 196GlyGlyAlaIleGlnThrSerAspAlaLeuThrIleThrGlyAsn 210
 3213 CCGCAGCAAGAAATGTACACCAAGAGACACATATCATCATGACACG 3262
 211 GlnGlyGlyValSerPheSerAspAsnThrSerAspSerGlyAlaI 227
 3263 ACAGCAGCAGGAGTAACTGCTCAAAATGTTACAAATTAAGGCAAC 3310
 227 aIlePheThrGluAlaSerValThrIleSerAsnAlaLysValSerP 244
 3311 ...ATTACCTCGCAAAATGTTACAGTACACCAACAAATCTTGTTA 3356
 244 heIleAspAsnLysValThrGlyAlaSerSerThr..... 256
 3357 CCACAGAGAAATGCTGCTAATGCAACCGCAGCAATTAACATTAAGT 3406
 257ThrGlyAspMetSerGlyGlyAlaIleGlyAlaThrLysTh 271
 3407 ACAAAACAGGGGATATTAAAGTGA.....ATTGAATCAACTTC 3447
 271 tThrAspThrLysValThrLeuThrGlyAsnGlnMetLeuPheSer 288
 3448 CGGTAATGTAATATTACAGGAGCGGCAAT...ACACTTAAGTAAGTA 3494
 288 snAsnThr.....Ser 291
 3495 ATATCACTGCTCAAGATGTACAGTACAGCGGATGACAGCCTTGACA 3544
 292 ThrThrAlaGlyGlyAlaIleThrValLys..... 301
 3545 ACTACAGAGGCTCAACCTTAATGTCGCAACAGGCAATGCAATATTAC 3594
 302LysLeuGluLeuAlaSerGlyL 310
 3595 AACCAAAACAGGTGATATCAACGGTAAGTGAATCCAGCTCCGCTGTG 3644
 310 euThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaProl 326
 3645 TAAACACTGTTGCAACTGAGCACTCTTGCTGTAGTAATATTTCAGGT 3694
 327 GlyAlaIleAlaIle...GluAspSerGlyGluLeu..... 337
 3695 AACACTGTTACTATTACTGCGGATAGCGTTAAATTAACTTCACACTG 3744
 338SerLeuSerAlaAspSerGly 345
 3745 TTCTACAAATTAATGAGCTAATAGTGAACCAACCAACCAATCAGGCG 3794
 345 spIle.....ValPheLeuGlyAsnThrValThrSerThrProGly 359
 3795 ATATTGAAGTACAAATTTCTGTATATACAGTAATATTGTTACA..... 3835
 360 ThrAsnArgSerSerIleAspLeuGlyThrSerAlaLysMetThrAla 376
 3836 GCAGCACTGCTGATTTAAGTATGGAATAGTCAAAAGTTGAAGCAA 3885
 376 uArgSerAlaIleGlyAlaIleThrPheThrAspProIleThrG 393
 3886 AAATGGAGCTCA..... 3898
 393 LysSerSerThrThrValThrAspValLeuLysValAsnGluThrProAla 409
 3898 3898

410 AspSerAlaLeuGlnIleThrGlyAsnIleIlePheThrGlyGlyL 426
 3898 3898
 426 uSerGluThrGluAlaIleAspSerLysAsnLeuThrSerLysLeuG 443
 3899ACCTTAAGTCTGATAGCAATTAATTAACCAACCAACGCT 3941
 443 InProValThrLeuSerGlyGly.....ThrLeuSerLeuLysHis... 456
 3942 CTACACTTACTCTCAAGCAATGCTGACAGCACTCTTACAGCAAGATAGC 3991
 457GlyValThrLeuGlnThrGln... 463
 3992 AGTATCCAGCAAAATTAATGCTGCTAATGTCAGTTAATACCAAGC 4041
 464 .AlaPheThrGlnGlnAlaAspSerArgLeuGluMetAspValGlyThr 479
 4042 CACTTAACTACTACAGGGGATTTAAAGATTAAACGCAACCAAGTGTACT 4091
 480 ..ThrLeuGluProAlaAsp.....ThrSerThrIleAsn 490
 4092 TAAACAATCAATGCAAAAGATGCCAAATTAGATGCTGCTCATCAGTGC 4141
 491 AsnLeuValIleAsnIleSerSerIleAspGlyAlaLysLysAlaLys 507
 4142 CGCAGCTAGTAATGCACTAACGCAAGTGGCTGTGTACGTG..... 4186
 507 eGluThrLysAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThr 524
 4187 ...ACAGCAAAACCTCAAGCAGCGGATATACCGCGGATTTAACA 4232
 524 euLeuAspProThrGlyThrPheThrGluAsn 534
 4233 CAATTAATGGTTAATATCATTCGCAAAAT 4264
 seq_name: /cgn2_6/ptodata/2/lna/5D_COMB.seq:US-08-617-697-8
 seq_documentation_block:
 ; Sequence 8, Application US/08617697
 ; Patent No. 5977336
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Matlare, Ltd.
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 5 Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/617,697
 ; FILING DATE: 01-APR-1996
 ; CLASSIFICATION: 424
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302,832
 ; FILING DATE: 05-OCT-1994
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651


```

1524 TCACAGCGCTGTGACGATCAGCATCTGCGCGCCCTCGCCACGAC 1573
      ::::::::::::::::::::
241 ..... LysValSerPheIleAspAsnLysValThrGlyAla..... 252
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1574 GCGCGCAAGATCGCCAGCGTCACTCGGCGACCTTGGCGCGCGCACGAT 1633
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
253 ..... SerSerSerThrThrGlyAspMetSerGly..... 262
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1624 CAGATGACCGCTCTGACGACCGTCAACCTGTGCGGACGCGCACCTGCG 1673
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
263 ..... GlyAlaIleCysAlaTyrLysThrSerThrAspThr 274
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1674 TCGGCATCGCGCGCGCTCTGACCGC..... 1702
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
275 LysValThrLeuThrGlyAsnGlnMetLeuLeuPheSerAspThrSer 291
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1703 ..... ACCGCGACCGCCACACCCCTGACCTTAACGTCAATGGTCTGAC 1746
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
291 rThrThrAlaGlyGlyAlaIleTyrValLysLysLeuGlnLeuAlaSer 308
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1747 GACGAC..... GCGCGATCAGGACCTGGAAGCGCTGCTGACGATG 1790
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
308 LysLeuThrLeuPheSerArgAsnSerValAsnGlyGlyThrAlaPro 324
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1791 GT..... TTCACCCACATCAACATCGTGTGTCGACCGCCTCT 1828
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
325 LysGly... GlyAlaIleAlaIleGlyAspSerGlyGlnLeuSerLeu 340
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1829 TCAGCATGCGCCAGCTGTGTGCGCGCGACGACGACACCCCTAATCATC 1878
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
340 rAlaAspSerGlyAspIleVal..... PheLeuG 350
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1879 GGGCGACGCTCCGCTCAGCATCCTGCGACACCGCTGCCCGCTGACGG 1928
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
350 LysAsnThrValThrSerThr..... 356
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1929 GCATCAGCGTGACCAACAGCGTGTGCGACCTCGCGCCGCACTGGCG 1978
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
357 ..... ThrProGlyThrAsnArgSerIleAspLeuGln 368
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1979 ACCGCTGTGCTCTCAGCGCGCGCTGCGCGTCACTGATCTCTGTGG 2028
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
368 YThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyAlaIleT 385
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2029 CCGCAGACGACAG..... GCGATCGTCAAGGGCGCGCGC..... 2062
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
385 YrPheTyrAspProIleThrThrGlySerSerThrVal..... 398
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2063 ..... GACGACACCGTCACTGCTGAGCTGGGACCTGGCGCTGCT 2104
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
399 ..... ThrAspValLeuLysValAsn..... 405
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2105 GGTTCGTCACGCGCGGACGCGACGACGAGCTGTCTGTGGCCAAAGTCAA 2154
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
406 ..... GluThr..... 407
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2155 CGGTTCGTCGTACAGCGCTGACCGGCTTCGCGGCTTCGAACCTTCC 2204
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
408 ... ProAlaAspSerAlaLeuGlnTyrThrGlyAsnIleLe... PheThr 422
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2205 GCGTCGCTGGCGCGGCTCAAGGCTCGCAACACCGCAACGCTTACG 2254
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
423 GlyGlnLysLeuSerGluThrGlnAlaAla..... 432
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2255 GCTCTGCACTGGCGCGACGCGGGTGCAGACGACTTCACCAACGTTGC 2304
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
432 ..... 432
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2305 GGTGAATGTGCGCTGACCGTTCGTGGGAGCTCGACCGGTACGAGACG 2354
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
433 ..... AspSerLysAsnLeuThrSer 439
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

2355 TGACCTTGCCACACCGCCAGCGGCACTCGGACGCTGTCAACCTGAC... 2401
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
440 LysLeuLeuGlnProValThrLeuSerGlyGlyThrLeuSerLeu..... 454
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2402 ... CTGTGCTCTGGCGCTCTGCGCTGTGAGGTTGCTGTGCTGCG 2448
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
455 ..... LysHisG 457
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2449 CCGTCAGCGGTGACATCGCGCCACCGACACCAACGACCGCCCTCAG 2498
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
457 LysValThrLeuGlnThrGlnAla..... PheThr 466
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2499 TCGACAGCTGACGCTGCAAGCCACCTCGCCAGATGATGATGTGAGAG 2548
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
467 GlnGlnAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuLys 483
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2549 GGCACGCGGCTGTGACCTGTACACACACCGGACACGCGCTGTACAG 2598
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
483 rAlaAspThrSerThrIle..... 489
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2599 CTTCGACGCGCCAGCCCGTCAACGCGACGCGCTCGCGTGTGACCTTGT 2648
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
490 ..... AsnAsnLeuValIleAsnIleSerSerIleAspGlyAlaLys 504
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2649 CCGCCACACACGCGGTGGTGAAGTGTACAGATCCGCGGCGCGCTGCG 2698
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
505 AlaLysIleGlnThrLysAlaThrSerLysAsnLeuThrLeu..... 518
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2699 GCGACGCTGTACCGGCTGCGGCGACCGCCAAATGACACATCATCGGTGG 2748
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
519 SerGlyThrIleThrLeuLeuAspProThrGlyThrPheTyrGlnAsn 535
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2749 CCGTGGCTACACCTGTGTACACCGCGGCTACG.....G 2786
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
535 LysSerLeuArgAsnProGlnSerTyrAspIleLeuGlnLeuLysAlaSer 551
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2787 ACACCTTACGCGGTGGCGAGCGGCGGATATCTTGATATCAACGCTATC 2836
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
552 GlyThrValThrSerThrAlaValThrProAspProIleMetGlyLys 568
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2837 GCGACCTGACCGCTTTCGTCGACGATCACCGACCGCTGTGCGCCACAA 2886
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
568 sPheHisTyrGlyTyrGlnGlyThrTyrGlyProIleValTyrGlyThr 584
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2887 GCTGACCTGTGCGCATCTGCGACGACGCGGTATGCTGACGCGCGCT 2936
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
585 ... GlyAlaSerThrThr 589
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2937 TCGCGCTGCGGTACCC 2953
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

seq_name: /cgn2_6/ptodata/2/ina/5D_COMB.seq:us-08-614-377A-6
seq_documentation_block:
; Sequence 6, Application US/08614377A
; Patent No. 5976864
; GENERAL INFORMATION:
; APPLICANT: Smith, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 5976864ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
; TITLE OF INVENTION: HEREROLOGOUS
; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson PC
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



```

385 yrhethyrasprollethrhrclysersethrthval..... 398
      |||
2063 .....SAGCAGACCGCTGACGCTCGGCGGCGCTGCT 2104
399 .....ThrAspValLeuLysValAsn..... 405
2105 GGTTCGTGACAGCGGCGGAGCGGACCGAGCTTCTGTGCGCCACAGCTCAA 2154
406 .....GluThr..... 407
2155 CGGTTCGTGCTCAGCGCTGACCGGCTTCGGCGGCTTCGAACCCCTCC 2204
408 ..ProLAspSerAlaLeuGlnThrLysGlnLelle...PheThr 422
      |||
2205 GCGTCGCTGGCGGCGGCTCAGAGCTCGACACAGCCACAGCGCTTCACG 2254
423 GlyGluLysLeuSerGluThrGluAlaLla..... 432
2255 GCTTCGACACTGGCGCGGCGGAGTGGACGACCTTCACCAACGTTGC 2304
432 ..... 432
2305 GGTGAATGTGGCTGACCGCTTCGGCGGCTCCGACCGGTACAGACCG 2354
433 .....AspSerLysAsnLeuThrSer 439
2355 TGACCTGTGCGACACCGCAGCGGACCTCGAGCGTGTTCACACTGAC... 2401
440 LysLeuLeuGlnProValThrLeuSerGlyLysThrLeuSerLeu..... 454
      |||
2402 ...CTGTGCTCTCGCGCGCTCGCGCGTGTAGGTTGGCTGCGTGGCG 2448
455 .....LysHisG 457
2449 CGTCGAGAGCGTGAAATCGCCGCCACCGACACACACAGACCGCTCCAG 2498
457 LysValThrLeuGlnThrGlnAla.....PheThr 466
2499 TCGACAGCGTGTGAGCTGACAGCCACTCGGCGCAAGTCGATCGTGTGACG 2548
467 GlnGlnAlaAspSerArgLeuGlnMetAspValGlyThrLeuGluPr 483
2549 GGCACAGCGCGGTGTGACCTGACCAACACGCGACACGCGTGTACACG 2598
483 oAlaAspThrSerThrIle..... 489
2599 CTTCGACGCGCAGCGCGCTCACCAGCGCGCTCCGGCTGTGACCTCGTGT 2648
490 .....AspAsnLeuValIleAsnIleSerSerIleAspGlyAlaLys 504
2649 CGGCCACACACAGCGTGGGTGAATGTCACATCCGCGCGGCGCTGCG 2698
505 AlaLysIleGluThrLysAlaThrSerLysAsnLeuThrLeu..... 518
2699 GCGGACTCGCTACCGGTTGCGCCACCGCAATGACACCAATATGCGTGG 2748
519 ..SerGlyThrIleThrLeuLeuAspProThrGlyThrPheYrgLysAsn 535
2749 CGGTGCGCTGACACCGCTGTCTACACCGCGGTACG.....G 2786
535 IserLeuArgAsnProGlnIserThrAspIleLeuGlnLeuLysLaser 551
2787 ACACCTTCACGGGTGGCAGCGCGGATGTCATTCATTCACACCTTC 2836
552 GlyThrValThrSerThrAlaValThrProAspProIleMetGlyGly 568
2837 GGCACCTGACCGCTTCGTGATCAGTACACCGCGCTGCGGAGAA 2886
568 sPheHisLysGlyYrgLysGlnGlyThrPrpGlyProIleValThr 584
2887 GCTCGACCTGCTGCGCATCTGACGAAACGCGCTATGCTGACGCGCGCT 2936

```

```

585 ..GlyAlaSerThrThr 589
      |||
2937 TCGGCGCTGGCGTACC 2953

```

```
seq_name: /cgn2_6/prodata/2/ina/5A_COMB.seq:US-08-038-682-5
```

```
seq_documentation_block:
```

```
; Sequence 5, Application US/08038682
; Patent No. 5549897
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: BARENKAMP, STEPHEN J
; TITLE OF INVENTION: ST. GEME III, JOSEPH W
```

```
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
```

```
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSER: Shoemaker and Mataré, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
```

```
; CITY: Arlington
; STATE: Virginia
```

```
; COUNTRY: U.S.A.
; ZIP: 22202-0286
```

```
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
```

```
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
```

```
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
```

```
; TELEFAX: (703) 415-0813
```

```
; INFORMATION FOR SEQ ID NO: 5:
```

```
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 9171 base pairs
```

```
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: DNA (genomic)
```

```
; US-08-038-682-5
```

```
alignment_scores:
```

```
Quality: 241.50 Length: 614
```

```
Ratio: 0.736 Gaps: 28
```

```
Percent Similarity: 53.420 Percent Identity: 22.638
```

```
alignment_block:
```

```
US-09-428-122-2 x US-08-038-682-5 ..
```

```
Align seg 1/1 to: US-08-038-682-5 from: 1 to: 9171
```

```
20 MetIleAlaThrGlnThrValLeuAspSerSerIleAspGlyAs 36
```

```
7752 ATGATGTGTAAGCATGTGTACCAAAAAAACAATACCTTGAAGGAGG 2801
```

```
36 nlyAsnGlyAsnPheserValArgLysGlnGlnIleAspAlaGlyThr 53
```

```
2802 TNAAGTG...AGGTTGGCTCCAGAAAGCGTACAGAA..... 2838
```

```
53 hrTYrLeuPheLysGlyAsnValThrLeuGlnAsnIleProGlyThrGly 69
```

```
2839 .....ATCGAAGCAATGTACTATCAATAAC.....AACGCTAAC 2874
```

```
70 ThrAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspLeuThr.. 85
```

2875 GTCACTCTTATCGGTTCGATTTCACACCATCAAAACCTTTAACTAT 2924
86PheThrGlyAsnGlyAsn 92
2925 TAAAAAGATGTCATCATTAATAGCGCAACCTTACCGCTGGAGGCAATA 2974
92 eLeuLeuPheGlnThrValAspAlaGlyThrValAlaGly..AlaAla 107
2975 TTGTC.....AATATACCGCGAAATCTTAC 3000
108 ValAsnSerSerValValAspLysSerThrThrPheIleGlyPheSer 124
3001 GTTGAAGATGAGCTTATTTCAAGCTATCACAAATTCACCTTTAATGT 3050
124 rLeuSerPheIleAlaSerProGly...SerSerIleThrGlyLysG 140
3051 AGCGCGCTGTGTGACACAAAGCAATTCAAATATTCATTGCCAAG 3100
140 lValAlaValSerCys.....SerThrGlySerLeuSerLeuThr 152
3101 GAGGGGCTCGCTTTAAGACATTGATTAATCCAGAAATTTAAGCATCAC 3150
153 LysAsnValSer.....LeuLeuPheSerLysAsnPheSerThr 165
3151 ACCAATCCAGCTCCACTACCGCACTATTATAGCGCAATATAACCA 3200
165 rAspAsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThr.....G 180
3201 TAAAAACGCT.....GATTAAATATTACGAACGAG 3232
180 lYThrThrMetSerAlaLeuPheSerGluAsnThrSerSerLysGly 196
3233 GTAGGATGTCGAATGCAATGCGGCGGATGTCGCCAAAAGAGGT 3282
197 GlyAlaIleGlnThrSerAspAlaLeuThrIleThr..... 208
3283 AATCTCAGCATTTCTTGCAAAATCAATATTACCAACAGATTAACAT 3332
209GlyAsnGlnGlyLysValSerPheSerAspAsnThrSerSer 223
3333 CAAGCGAGGTGTGATGGGAGAAATTCGATTCCAGCGCACAAACAAATG 3382
223 sPserGlyAlaAlaIlePheThrGluAlaSerValThr..... 235
3383 CCAATCTAACCATTAACCAAGAAATGAATTAACGACAGACTTAAT 3432
236 lIleSer.....AsnAsnAlaLysValSerPheIleAspAsnLysValTh 250
3433 ATTTCAAGTTTCATATAAGCAGAGATTACAGCTAAAGATGTAGTGAATT 3482
250 rGlyAlaSerSerSerThrThrGlyAspMetSerGlyGlyAlaIleCys 267
3483 AACTTTGTGTAACCAATAGTGTGATGTAATGCCAAAAGATTA 3532
267 lATyrlYThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMet 283
3533 CCTTAACAGGTTAAAGATTCAAAAATCTCTGCGAGCGGTCAAGG 3582
284 LeuLeuPheSer.....AsnAsnThrSerTh 292
3583 ACATTAACAGCAAAAGTGCAACATCCGGTACTAATACCAACAGCTGA 3632
292 rThrAlaGlyGlyAlaIleTyValLysLysLeuGluLeuAlaSerGly 309
3633 TAGCAGTGCAAT.....AATGCCG 3652
309 lYLeuThrLeuPheSerArgAsn...SerValAsnGlyGlyThrAlaPro 324
3653 GCTTAACATTCATGCAAAAATGTAACAGTAACAAACAAAT..... 3693
325 LysGlyGlyAlaIleAlaIleGluAspSerGlyLysLeuSerLeuSerAl 341
3694ATTACTTCTCAAAAGAGTGAGCATCTCTGC 3725

341 aAspSerGlyAspIleValPhe...LeuGlyAsnThrValThrSerThr 357
3726 GACAGGTGGAATAATTACACTAAACAGTACAAACCATTAACCAACCA 3775
357 hrrProGly.....ThrAsnArgSerSerIleAspLeuGly..... 368
3776 CTGTAAACGTGAGATTAACCGCTCAAAACAGTACTATCTGTAGTGGAAT 3825
369ThrSerAlaLysMetThrAlaLeuArgSerAlaAlaGlyArgAl 383
3826 GAGTCCAGCTCTGGCTGTGAACACTTACTGCAACCGAGGCGCTCTGC 3875
383 aIleTyrlPheTyrlAspProIleThrThrGlySerSerThrThrValThra 400
3876 TGT.....ACCAATATTTCGCGCAACACCTGTCTGTACTG 3913
400 sPvalLeuLysValAsnGlnThrProAlaAspSerAlaLeuGlnTyThr 416
3914 CAAATAGCGGTGATTAACCACTTTGGCAGGCTCTACAAATTAAGAAC 3963
417 GlyAsnIleIlePheThrGlyLysLeuSerGlnThrGluAlaAlaAs 433
3964 GAGACTGA.....ACCACTTCAAGTCA 3986
433 pSerLysAsnLeuThrSerLysLeuLeuGlnProValThrLeuSerGly 450
3987 ATCAGCGCATATCGCGGT.....ACGATTTCTGTGTG 4018
450 lYThrLeuSerLeuLysIleGlyValThrLeuGlnThrGluAlaPheThr 466
4019 GCACAGTAGAGTTAA.....GCAACCGAAAGTTAAC 4053
467 GlnGluAlaAspSerArgLeuGlnMetAspValGlyThrThrLeuGluPr 483
4054 ACTCATTCATTCAAAATTAAGCAACACAGCGGCTAACGTAAAC 4103
483 oAlaAspThrSerThrIleAsn.....AsnL 492
4104 AAGTGCAACAGGTACAAATTTGGTGTAGCATTTCCGGTAATCGGTAAATG 4153
492 euValIleAsn.....IleSer 497
4154 TTACGGCAAAACGTGGCGATTTAACAGTTGGGAATGGCGCAAAATTAAT 4203
498 SerIleAspGlyAla.....LysLysAlaLysIleGlnTh 509
4204 GCGACAGAGAGCTGCACCTTAACATCAATCGGGCAAAATTAACATAC 4253
509 rLysAlaThrSerLysAsnLeuThrLeuSerGlyThrIleThrLeuLeu 526
4254 CGAAGCTAGTTCACACATTAATCAAGCAAGGTCAAGTAATCTTTCAG 4303
526 sProThrGlyThrPheTyrlGluAsnHisSerLeuArgAsnProGlnSer 542
4304 CTCAGAGATGAGCGTTCAGAGAGTATTAATGCCGCAAT..... 4344
543 TyrAspIleLeuGluLeuLysAlaSerGlyThrValThrSer 556
4345GTGACACTAAATATCAAGGCACTTAACCTAC 4377

Fri Nov 24 13:50:08 2000

us-09-428-122-2.rni

Page 24

OM of: US-09-428-122-2 to: EST.* out_format: pfs

Date: Nov 20, 2000 3:57 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+pn.model -DEV=xlp
-O/cgnt2.1/USPTO.spool/US09428122/runat.14112000.163316.27900/app_query.fasta.1.999
-DB=EST -OFMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-TEAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blissum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=400000000 -USER=US09428122.@CGN1.1.2757 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPPX -WAIT -THREADS=1

Search information block:

Query: US-09-428-122-2

Query length: 928

Database sequences: 7189864

Database length: -1203564053

Search time (sec): 6294.330000

Score list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_gss23:CNS04ALV	171.00	304.40	5.4e-08	898	AL282028 Tetraodon nigroviridis
gb_gss23:CNS04RM2	158.50	281.26	1.1e-06	805	AL304427 Tetraodon nigroviridis
gb_est15:AV062062	140.00	246.19	9.5e-05	730	AU062062 A062062 Dictyostelium
gb_gss21:CNS00GFX	131.50	225.33	0.0014	1006	AL0072139 Drosophila melanogaster
gb_gss12:AO875306	129.50	229.95	0.0008	520	AO875306 V12294 mtu-3xHA/lacZ
gb_est11:AM256683	127.50	221.66	0.0022	727	AM256683 EST304820 KV2 Medicago
gb_gss9:AO645402	127.00	221.70	0.0022	673	AO645402 RPI93-Dpmlt-29D12.T9
gb_est18:AV403135	126.00	223.38	0.0017	500	AV403135 AV403135 Bombyx mori
gb_est18:AV403126	126.00	220.19	0.0027	650	AV403126 AV403126 Bombyx mori
gb_est11:AI151569	125.00	217.39	0.0038	693	AI151569 xylem.est.507 Poplar x
gb_gss24:CNS05GDN	124.00	217.02	0.0040	613	AI151569 GH28545.5prime GH Dros
gb_gss24:CNS05GDN	123.50	218.84	0.0115	1068	AI151569 GH28545.5prime GH Dros
gb_gss9:AO640209	123.00	215.79	0.0047	579	AO640209 927P1-2P1.TV 927P1.TV
gb_est8:AI115059	121.00	207.35	0.0139	820	AI115059 ui41h04.y1 Sugaano mos
gb_gss23:CNS03T7	121.00	205.89	0.0167	918	AL259756 Tetraodon nigroviridis
gb_est1:AF031703	121.00	202.99	0.0243	1149	AF031703 AF031703 Human lymph
gb_est25:AM906712	120.50	207.20	0.0141	769	AM906712 EST342835 potato stoid
gb_gss9:AO639357	120.00	208.68	0.0117	655	AO639357 927P1-10F10.TV 927P1.T
gb_est14:BE247955	120.00	208.30	0.0123	655	BE247955 NF002D12P1F1094 Dros
gb_est15:AU061810	120.00	207.15	0.0142	716	AU061810 AU061810 Dictyostelium
gb_gss23:CNS04TTL	120.00	204.61	0.0197	871	AL306930 Tetraodon nigroviridis
gb_est11:AI151361	119.50	206.33	0.0154	696	AI151361 GH28545.5prime GH Dros
gb_est11:AM256769	119.50	206.36	0.0157	705	AM256769 EST304906 KV2 Medicago
gb_est16:C93043	119.50	205.77	0.0170	738	C93043 C93043 Dictyostelium dis
gb_est13:BE130661	119.50	205.26	0.0182	768	BE130661 L48-86273 Ice plant La
gb_est13:BE034472	119.50	205.22	0.0182	770	BE034472 MH05E02.MH Mesembryant
gb_est13:BE130655	119.50	205.10	0.0185	777	BE130655 L48-86273 Ice plant La
gb_est13:BE131175	119.50	204.91	0.0190	789	BE131175 L48-113073 Ice plant La
gb_est13:BE033730	119.50	203.69	0.0222	867	BE033730 MF07F05.MF Mesembryant
gb_est13:BE033730	119.50	202.26	0.0267	968	BE033730 MO20D11 MO Mesembryant
gb_est13:BE033730	119.50	201.92	0.0279	995	BE033730 MO10C04 MO Mesembryant
gb_gss23:CNS04NSM	119.00	204.84	0.0191	735	AL2693119 Tetraodon nigroviridis
gb_est23:AM640354	119.00	203.58	0.0225	810	AM640354 b194c08.w1 Blackshear
gb_gss22:CNS02PLH	119.00	201.92	0.0278	921	AL208142 Tetraodon nigroviridis
gb_gss23:CNS03XRI	119.00	199.62	0.0340	1100	AL265583 Tetraodon nigroviridis
gb_est14:AU004371	118.50	203.07	0.0240	781	AU004371 AU004371 Bombyx mori
gb_est13:BE033623	118.50	199.52	0.0379	1028	BE033623 MO12D08 MO Mesembryant
gb_est13:AM052690	118.00	207.14	0.0143	529	AM052690 rs38e02.y1 Sommer pris
gb_gss24:FR003380	118.00	205.14	0.0184	617	AL030167 Fugu rubripes GSS seq
gb_gss21:CNS00EDS	118.00	197.65	0.0482	1101	AL068864 Drosophila melanogast
gb_gss22:CNS03151	117.50	199.58	0.0376	879	AL223102 Tetraodon nigroviridis

gb_gss17:AZ163905 - 117.00 205.25 0.0182 526 ! AZ163905 SP_0074_B2_B08_T7A
gb_gss12:AO855002 + 117.00 199.29 0.0390 833 ! AO855002 CpG2052B CpIOMAGDNA
gb_est11:AI534939 + 116.50 203.06 0.0240 577 ! AI534939 SD01229.5prime SD D
gb_est14:BE253867 + 116.50 195.12 0.0666 1066 ! BE253867 601112837F1 NIH_MG

seq_name: gb_gss23:CNS04ALV

seq_documentation_block:

LOCUS CNS04ALV 898 bp DNA 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
095820 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL282028
VERSION AL282028.1 GI:8020362
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
Holacanthopterygii; Acanthopterygii; Perciformes;
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.
1 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 898)
Direct Submision
Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

REFERENCE
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 898)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and
Weissenbach,J.<

```

168 IYGLYAlaIleThrAlaIysThrLeuSerLeuThrGlyThrThrMetSer 184
      ::::: |||::: ||| |||::: |||
112 AAGGTTCTTACTGTTCTGGAAGCTTCTGACAGCTTCTAAAGTTCT 161
      ::::: |||::: ||| |||::: |||
185 AlaLeuPheSerGluAsnThrSerSerLysGlyAlaIleGln 201
      ::::: |||::: ||| |||::: |||
162 TCTACTGTTCTAAAGTTCTCT ..... 185
201 rSerSpAlaLeuThrIleThrGlyAsnGlyGluValSerPheSerA 218
      ||| |||::: |||::: |||::: |||
186 .....ACTGTTCTGAAGCTTCTTCAACAGCTTCTA 216
218 SpASnThrSerSerAspSerGlyAlaAlaIlePheThrGluAlaSerVal 234
      ::::: |||::: ||| |||::: |||
217 AAGGTTCTTACTGTTCTGGAAGCTTCTGACAGCTTCTAAAGTTCT 266
      ::::: |||::: ||| |||::: |||
235 ThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysValThrG1 251
      ::::: |||::: ||| |||::: |||
267 ACAGGTTCTTAATGTTCTTCTACTGTTCTGAAGCTTCTCTATTT...GG 313
      ::::: |||::: ||| |||::: |||
251 YAlaSerSerSerThrThrGlyAspMetSerGlyAlaIleGlyAlaIle 268
      ::::: |||::: ||| |||::: |||
314 TTCTGAAGCTTCTTCTACTGTTCTAAAGT ..... 344
266 YrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGlnMetLeu 284
      ::::: |||::: ||| |||::: |||
345 .....TCTTATGTTGTTCTGGAAGCTTCTTCTACTGTTCTGGAAGCTTTC 389
285 LeuPheSerAsnAsnThrSerThrThrAlaGlyAlaIleThrValLys 301
      ::::: |||::: ||| |||::: |||
390 TCTACTGTTCTGAAGCTTCTTCTACTGTTCTAAAGCTTCTCTAATG 439
301 sLysLeuGluLeuAlaSerGlyLeuThrLeuPheSerAspAsnSerVal 318
      ::::: |||::: ||| |||::: |||
440 TTCT...GAAGCTTCTTACTGTTCTGGAAGCTTCTTCTACTGTTCTG 486
318 AlaSerGlyThrAlaProLysGlyAlaIleAlaIleGlnAspSer 334
      ::::: |||::: ||| |||::: |||
487 AAGGTTCTTCAAGGCTTCTAAAGCTTCTTATGTTCTTAAAGCTCT 536
335 GlyLeuLeuSerLeuSerAlaAspSerGlyAspIleValPheLeuGlyAs 351
      ::::: |||::: ||| |||::: |||
537 TCCATGTTCTGAAGCTTCTTCTAATGT ..... 566
351 nThrValThrSerThrThrProGlyThrAspAspSerSerIleAspLeu 368
      ::::: |||::: ||| |||::: |||
567 .....TCTAAGCTTCTTCTACTGTTCTGAAATCTTCTCTACTG 606
368 IYThrSerAlaLysMetThrAlaLeuArg...SerAlaAlaGlyArgAla 383
      ::::: |||::: ||| |||::: |||
607 GTTCTAAAGTTCTTCCCTGTTTACATTTTCTACTTTTCTTTTGT 656
384 IleThrPheThrAspProIleThrThrGlySer.....SerThrThrVa 398
      ::::: |||::: ||| |||::: |||
657 TCTGAAGCTTCAAGTGTCTTCTACTGTTCTGTAAGCTTCTTCAAGCG 706
398 LThrAspValLeuLysValAsnGluThrProAlaAspSerAlaLeuGln 415
      ::::: |||::: ||| |||::: |||
707 TTCTGAAGCTTCTACTGTTCTGAAACTTCTGATGTTCTCTC..... 749
415 YrThrGlyAsnIleIlePheThrGlyLysLeuSerGluThrGluAla 431
      ::::: |||::: ||| |||::: |||
750 .....ACTGTTCTTAAGTTCTTCTTCTACTGTTCT 779
432 AlaAspSerLysAsnLeuThrSerLysLeuGlnProValThrLysLeu 448
      ::::: |||::: ||| |||::: |||
780 GAGACCTTACTGTTTGAAGCTTCTTCTGCAATCTAAAGCTTCTTC 829
448 rGlyGlyThr 451
      ::::: |||::: ||| |||::: |||
830 TTCTGTTCT 839

```

```

seq_name: gb_gsa23:CNS04RM2
seq_documentation_block: 805 bp DNA GSS 24-MAY-2000
LOCUS CNS04RM2 tetraodon nigroviridis genome survey sequence PUC-Orig end of clone
DEFINITION 005020 of library H from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL304427.1 GI:8192034
VERSION AL304427.1
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
Holacanthopterygii; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE Roest-Crollius,H., Jallion,O., Dasilva,C., Fizes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 805)
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizes,C., Mincker,P., Brotlier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
COMMENT Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
3 (bases 1 to 805)
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 805)
AUTHORS Genoscope
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1..805
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="005020"
/clone_lib="H"
/note="Genoscope sequence ID : COBH005B10XE1-end ;
PUC-Orig"
BASE COUNT 87 a 194 c 293 g 226 t 5 others
ORIGIN
alignment_scores:
Quality: 158 50 Length: 211
Ratio: 1.101 Gaps: 4
Percent Similarity: 66.246 Percent Identity: 27.014
alignment_block:
US-09-428-122-2 x CNS04RM2/rev ..
Align seg 1/1 to reverse of: CNS04RM2 from: 1 to: 805
117 ThrThrPheIleGlyPheSerSerLeuSerPheIleAlaIleSerProGlySer 133
      ::::: |||::: ||| |||::: |||
800 AGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 751
133 rSerIleThrThrGlyGlyGlyAlaValSerCysSerThrGlySerLeu 150
      ||| |||::: ||| ||| |||::: |||
750 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 701
150 rLeuThrLysAsnValSerLeuLeuPheSerThrLysAspPheSerThrAsp 166
      ||| |||::: ||| ||| |||::: |||
700 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 660

```



```
167 AsnGlyGlyAlaIleThrAlaLysThrLeuSerLeuThrGlyThrThrMe 183
      ::::::::::::::::::::||| ||| ::::::::::::::::::::
659 AGCAGCACACACACACAGCAGCAGCAGTGCAGCAGCAGCAGCAGCAGCAGC 610
183 tSerAlaLeuPheSerGlyAsnThrSerSerLysGlyAlaIleGly 200
      ||| ::::::::::::::::::::||| ::::::::::::::::::::
609 CAGTAGC.....AGCAGCACACACAGCAGCAGCAGCAGCAGCAGCAGCAG 566
200 LntHrSerAspAlaLeuThrIleThrGlyAsnGlyGlyValSerPhe 216
      ::::::::::::::::::::||| ||| ::::::::::::::::::::
565 GCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 519
217 SerAspAsnThrSerSerAspSerGlyAlaIlePheThrGlyAlaSer 233
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
518 AGCAGCACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 469
233 tValThrIleSerAsnAlaLysValSerPheIleAspAsnLysValT 250
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
468 CACAGCAGTGCAGCAGCAGTACAGCAGCAGCAGCAGCAGCAGTACAGCA 419
250 hrGlyAlaSerSerSerThrThrGlyAspMetSerGlyAlaIleCys 266
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
418 CCACCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTAC 369
267 AlaTyrLysThrSerThrAspThrLysValThrLeuThrGlyAsnGly 283
      ::::::::::::::::::::||| ||| ::::::::::::::::::::
368 AGCAGCACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 319
283 tLeuLeuPheSerAsnAsnThrSerThrThrAlaGlyAlaIleTyrIv 300
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
318 CACAGCAGTGCAGCAGCAGTACAGCAGCAGCAGCAGCAGCAGCAGTCCGA 269
300 aLysLysLeuGlyLeuAlaSerGlyLeuThr..... 311
      ::::::::::::::::::::||| ||| ::::::::::::::::::::
268 CCCAGGTCTTCCAGCAGCAGCAGCAGCAGTCTTCCAGTCCAGCAGCAGT 219
312 LeuPheSerArgAsnSerValAsnGlyGlyThr 322
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
218 CTGCTGCCCAATCATCCGCTGCAGCAGCAGCAGT 186
```

seq_name: gb_est15:AU062062

seq_documentation_block:

LOCUS AU062062 730 bp mRNA EST 20-MAY-1999
DEFINITION AU062062 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
discoideum cDNA clone SLH423, mRNA sequence.

ACCESSION AU062062
VERSION AU062062.1 GI:4883166

KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
Eukaryota; Dictyostelidae; Dictyostelium.

REFERENCE 1 (bases 1 to 730)
AUTHORS Yoshino,R., Morio,T. and Tanaka,Y.
TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1997)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
Email: d402h@tsukuba.ac.jp

PROJECT = Dictyostelium discoideum cDNA project in Japan.
Location/Qualifiers
1. 730

FEATURES
source
1. 730
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="SLH423"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"

BASE COUNT 247 a 137 c 150 g 196 t
ORIGIN

alignment_scores:
Quality: 140.00 Length: 286
Ratio: 0.915 Gaps: 8
Percent Similarity: 53.497 Percent Identity: 24.126
alignment_block:
US-09-428-122-2 x AU062062 ..

Align seq 1/1 to: AU062062 from: 1 to: 730

```
78 AsnAsnThrLysGlyAspLeuThrPheThrGlyAsnGlyAsnSerLeuLe 94
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
16 AATAGTGTCCGAGCAGATCAGATCAGTGCAGCAATAGTGTCAATATAC 65
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
94 uPheGlnThrValAspAlaGlyThrValAlaGlyAlaAlaValAsnSers 111
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
66 TGTATTACAACAATAGTGTGAATAGTGAAGTAAATGCTATC..... 108
111 ervaValAspLysSerThrThrPheIleGlyPheSerSerLeuSerPhe 127
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
109 .....GCAGCGCTAACCACTTCAAAATCA 132
128 IleAlaSerProGlySerSerIleThrThrGlyLysGlyAlaValSerCy 144
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
133 ATCCCTCAAACTGATCAGTACAGATTAATGCAATATGATTCAGCAG 182
144 sSerThrGlySerLeuSerLeuThrLysAsnValSerLeu..... 158
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
183 ATCAGTCGAGGTTTAAACACAGCAAACTCATTAACAGAGAGTTCA 232
159 .....PheSerLysAsnPheSerThrAspAsnGlyGlyAlaIleThr 172
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
233 TTGGTAATATCGCCAAATGCAATTTCACTGCTCTTGTGCTCAAAATCA 282
173 AlaLysThrLeuSerLeuThrGlyThrThrMetSerAlaLeuPheSerG 189
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
283 ACATCAAAACCAACAGTTTCATCACACAAACAACAATCA.....AGTCA 326
189 uAsnThrSerSerLysLysGlyAlaIleGlnThrSerAspAlaLeuT 206
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
327 AGGTACAAAGTGTATTATCTCTGAGCAGTACTATCTGCTGCTGATAT 376
206 hrIleThrGlyAsnGlyGlyValSerPheSerAspAsnThrSerSer 222
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
377 CAATT...GGTAATAGTGCATCAGTACCC...GGAATAACAATCAATCA 420
223 AspSerGlyAlaIlePheThrGlyAlaSerValThrIleSerAsnAs 239
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
421 GGAAGTTTACTGTGTCGAATACAGTACAGTACAGTACACAGCAACTC 470
239 nAlaLysValSerPheIleAspAsnLysValThrGlyAlaSerSerSert 256
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
471 A.....ATTGATCCAAACACCA 490
256 hrThrGlyAspMetSerGlyAlaIleCysAlaTyrLysThrSerThr 272
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
491 TAACTGTGTTTCATCATTCAGCAATTCACAGTTCAATGATGATTA 540
273 AspThrLysValThrLeuThrGlyAsnGlyMetLeuLeuPheSerAsn 289
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
541 ACAACTCAACATCATTAATGATATCATCAATGCTGCTGATTAATCA 590
289 nThrSerThrThrAlaGlyAlaIleTyrValLysLysLeuGlyLeuA 306
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
591 TCAAGCATCATCAATGCCAATGCAAT..... 618
306 laseGlyGlyLeuThrLeuPheSerArgAsnSerValAsnGlyGlyThr 322
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
618 ..... 618
323 AlaProLysGlyAlaIleAlaIleGlyAspSerGlyLeuLeuSerLe 339
      ||| ::::::::::::::::::::||| ||| ::::::::::::::::::::
```

```

619 .....TCAGGTGGTAATATTGACGGTGAACACTCACTCATAAATTCCAAAC 662
339 userAlaSpSerGlyAspIleValPheLeuGlyAsnThrValThrsert 356
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
663 CGGATATATATCGCA.....GTAATGCAGAGACATCTGCATCATCA 706
356 hrtHpro 358
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
707 CATCTCCA 714
seq_name: gb_gss21:CNS00GFY

seq_documentation_block:
LOCUS      CNS00GFY          1006 bp        DNA           GSS       03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence 17 end of BAC:
            BAC33119 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL072139
VERSION     AL072139.1   GI:4951919
KEYWORDS    fruit fly.
SOURCE      Drosophila melanogaster
ORGANISM    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachyoceta;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   Genoscope.
AUTHORS     Direct Submission
TITLE       Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL     BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT     - Web : www.genoscope.cns.fr)
             Determination of this BAC-end sequence was carried out as part of a
             collaboration with the Berkeley Drosophila Genome Project (BDGP).
             The BDGP is constructing a physical map of the Drosophila
             melanogaster genome using these BACs. For further information
             please see http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library.html
             A more detailed description of the library
             and how to order individual BAC clones, the entire library, or
             filters for hybridization from the BACPAC Resource Center can be
             found at http://bacpac.med.buffalo.edu/Drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..1006
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR33119"
                     /note="end : 17"
BASE COUNT         271 a      289 c      187 g      204 t      55 others
ORIGIN
Alignment_scores:
Ratio: 131.50      Length: 341
Percent Similarity: 57.185      Percent Identity: 22.287
Percent Simlarity: 57.185      Gaps: 13
alignment_block:
US-09-428-122-2 x CNS00GFY ..
Align seg 1/1 to: CNS00GFY from: 1 to: 1006
60 ValThrLeuGIuaSnIIeProGlyThrGlyThrAlaIleThrIlysSecry 76
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
21 ATGCTACCANAATCCCCGAAATATGCCCACTTGCGTCGAC 70
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
76 sphenasnhthrIysGlyAsnThrPheThrGlyAlaSnGlyAsnserL 93
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```



```

/clone_lib="KV2"
/tissue_type="Seeding roots"
/dev_stage="2 days post-inoculation with Sinorhizobium
melliotti"
/lab_host="E. coli strain SOLR"
/Note: Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The
cDNA was directionally ligated into the Unizap XR vector
from Stratagene and packaged using GigaPack III Gold
packaging extracts. Plasmids containing cDNA inserts
were excised from the recombinant lambda-Zap phage using
Ex-assist helper phage and propagated in SOLR cells."
BASE COUNT      294 a      125 c      172 g      136 t
ORIGIN

```

```

alignment_scores:
  Quality: 127.50      length: 305
  Ratio: 0.729        Gaps: 12
  Percent Similarity: 57.377      Percent Identity: 23.279

```

```

alignment_block:
US-09-428-122-2 x AM256683/rev ..

```

```

Align seg 1/1 to reverse of: AM256683 from: 1 to: 727

```

```

97 ThrValAspAlaGlyThrValAlaGlyAlaValAspSerValVa 113
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
727 ACTGTTCTGTAGGCTCACTCACTCACTCACTCACTCACTCACT 678
113 IAspLysSerThrThrPheIleGlyPheSerSerLeuSerPheIleAla 130
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
677 TTCTGTAGCTGAAGCTTCACTCACTCACTCACTCACTCACTT 631
130 exProGlySer...SerIleThrProGlyLysGlyAlaValSerLys 145
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
630 CTCTAGGTCACCTCACTCACTCACTCACTCACTCACTCACT 584
146 ThrGlySerLeuSerLeuThrLysAsnValSerLeuLeuPheSerLys 162
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
583 GTCAGATCACTCACTCACTCACTCACTCACTCACTCACTCACT 534
162 nPheSerThrAspAsnGlyAlaIleThrAlaLysThrLeuSerLeuT 179
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
533 AACTTCTGTAGCCACAGCTTCTTCTTCTTCTTCTTCTTCTT 496
179 hGlyThrThrMetSerAlaLeuPheSerGluAsnThrSerSerLys 195
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
495 TAGGATCACTCACTCACTCACTCACTCACTCACTCACTCACT 446
196 GlyAlaIleGlnThrSerAspAlaLeuThrIleThr...GlyAsnG 211
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
445 TCGAATTCAGTACGTCAGTTCATCTTCACTTCTGTAGGCTCAAC 396
211 nGlyAlaValSerPheSerAspAsnThrSerSerAspSerGlyAlaI 228
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
395 TAGAGTTCAGTACGTTCTTCTTCTTCTTCTTCTTCTTCTTCA 346
228 IepheThrGluAlaSerValThrIleSerAsnAsnAlaLysValSerP 244
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
345 TAGCCACAGANTCTCTTCTTCTTCTTCTTCTTCTTCTTCTT 317
245 IleAspAsnLysValThrGlyAlaSerSerSerThrGlyAspMetS 261
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
316 .....GTAGCTCAACTCACTGAGCCACAGCTTCT... 287
261 rGlyGlyAlaIleGlyAlaLysThrSerThrAspThrLysValThr 278
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
286 .....TCTTTACTGTTCTGTAGCTTCAACTTAGAG 253
278 euThrGlyAsnGlnMetLeuPheSerAsnAsnThrSerThrThrAla 294
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
252 CCACTGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 203

```

```

295 GlyAlaIleThrValLysLysLeuGluAlaSerGlyLeuThr 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
202 GGTTCTCTTTTACCGCTT.....TCTGTACCTCACTCACTGTC 162
311 rLeuPheSerArgAsnSerValAsnGlyThrAlaProLysGlyAla 328
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
161 CACAGGTTCTTCTTTTACCGCTTCTGTAGGCTCACT.....T 124
328 lAlaIleAlaGluAspSerGlyLysLeuSerLeuSerAlaAspSerGly 344
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
123 CACTACCCACAGGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCA 77
345 AspIleValPheLeuGlyAsnThrValThrSerThrProGlyThrAs 361
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
76 .....GTAGCTACAGCTTCTTCTTCTTCTTCTTCTTCTTCTT 53
361 nArgSerSerIleAspLeuGlyThrSerAlaLysMetThrAlaLeuArg 378
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
52 .....GTTTCTGTAGATCACTCA..... 32
378 eAlaIleAlaGlyArgAlaIleThrPheThrAspProIleThrThrGlySer 394
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
31 .....GTACACACAGCTTCTG 17
395 SerThrThrValThr 399
16 TCTTTACGCTTCT 2
seq_name: gb_gss9:AO645402
seq documentation block: 673 bp DNA GSS 08-JUL-1999
LOCUS AO645402
DEFINITION RC193-DpniI-29D12.1J RC193-DpniI Trypanosoma brucei genomic clone
ACCESSION RC193-DpniI-29D12.1J RC193-DpniI-29D12.1J RC193-DpniI-29D12.1J
VERSION AO645402.1 GI:5122112
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 673)
El-Sayed,N., Zhao,S., Zhao,H., Gull,S., Suh,E., Malek,J., Fujii,C.,
Gerriard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
Fraser,C. and Adams,M.
Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RC1-93
Library for gene discovery and sequence-ready map construction
Other GSSs: RC193-DpniI-29D12.TV
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
7712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@igrr.org
Clones and high density filters may be purchased from BACPAC
Resources (http://bacpac.med.buffalo.edu). BAC end sequences search
page: http://www.tigr.org/tdb/mdb/tbdb/.
seq primer: SP6
Class: BAC ends.
FEATURES
Source
Location/Qualifiers
1..673
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="RC193-DpniI-29D12"
/clone_lib="RC193-DpniI-29D12"
/Note="Vector: pBAC3.6; Site_1: Bam HI; Site_2: Bam HI;
Constructed for The Institute for Genomic Research by
Bohai Zhao in Pieter de Jong's laboratory (Roswell Park
Cancer Institute, Buffalo, NY). Briefly, Trypanosoma

```

[illegible]

seq_name: qb_est18:AV403135	
seq_documentation_block:	
LOCUS	500 bp
AV403135	mRNA
	EST
	06-FEB-2000

LOCUS	500 bp	mRNA	EST	06-FEB-2000
AV403135	Bombyx mori middle silk gland	5th-instar	larva	Bombyx mori
DEFINITION	AV403135 Bombyx mori middle silk gland 5th-instar larva Bombyx mori			
CDNA	clone msgv0085 T3, mRNA sequence.			
AV403135				
ACCESSION				

```

VERSION AV403135.1  GI:690
KEYWORDS EST

```

ORGANISM Bombyx mori
Eukaryota: Metazoa: Arthropoda: Tracheata: Hexanoda: Insecta:

Bombycoidea; Bombycidae; Bombyx.

[illegible]

COMMENT: MILD R
Genome Research Group

Ahagawa 4-9-1, Inage, Chiba 203-8555, Japan
Email: kmitta@ntrs.go.jp

(5' -> 3')
Project='Silkworm Genome Program in MAF, and Research for the

FEATURES
<http://www.ab.a.u-tokyo.ac.jp/silkbases/>, for whole ESTdb.
Location/Qualifiers

```
/organism="Bombyx mori"  
/db_xref="taxon:7091"
```

```
/clone_lib-"Bombyx mori middle silkgland 5th-instar larva"  
new #female /cell "mixed"
```

```

/issue_type="invertebrate larva"
/dev_stage="5th-instar larva"

```

BASE COUNT	142 a	129 c	123 g	105 e	100 f
ORIGIN					

alignment_scores:

Ratio:	1.156	Gaps:	7
Percent Similarity:	56.771	Percent Identity:	28.125

alignment_block:
MS-09-428-122-2 X AV403135

Align seg 1/1 to: AV403135 from: 1 to: 500

[illegible]

125 userpheilealaserproglyserSerileThrThrGlyGlyAlay 142

[illegible]

Fri Nov 24 13:50:17 2000

us-09-428-122-2.rst

Page 8

```
142 a1sercysertThrglySerleuserleuthrlyAsnValserleu 158
||||| ||||| ||||| |||||
116 TATCA...TCACCGCGCAGTTCACGATACACT..... 144
159 PheserlyAsnPheserThraspanglyAlaIlethrlalysrth 175
||| ||||| ||||| |||||
145 .....GATGACAGACAGACCTTCAGATTCACATCGGCGG 185
175 rleuserleuthrlyThrmetservalaleupheserlyAsnThr 192
||| ||| ||| |||
186 TACACCACTTATGGATACAGTTCACAGCT....CGTATGAGAGAG 229
192 erserlyslgylAlaIleGlnThrsaspAlaIlethrllethr 208
||| ||||| ||||| |||||
230 TATCATCCACGCGCAGTTCACATGATGACAGACAGACCTTACA 279
209 GlyAsnGln.....GlyGluValserPheserAsnThrSers 222
||| ||||| ||||| |||||
280 GGATCCAGTACATCCGCGGTACACACTATGATATAGTCACAGCA 329
222 rAspserGlyAlaIlePheThrGluAlaSerValThrIleSers 239
||| ||||| ||||| |||||
330 TCGTATGAGAGTATGTGCGCAGCTGCGAGT.....TCCAGTA 367
239 snAlaIlyValserPheIleAspAsnlyValThrglyAlaSer 255
||| ||||| ||||| |||||
368 ACACGTATGACAGACACAGAGATCCACACGTCGCTGATACAGC 417
256 ThrThrglyAspMetserGlyAlaIleCysAlaIlyThrSert 272
||| ||| ||| |||
418 ACTGAGGATATAGTTCC.....AGTAGCA 443
272 rAspThrlyValThrleuthrGlyAsnGlnMetleuPheSers 289
||| ||||| ||||| |||||
444 TGATGAGAGCTTACA.....TCCACCG 466
289 snrThserThrAlaGlyAla 297
||| ||||| ||||| |||||
467 ACGGTCCAGCACAGTGGAGAGCT 492
seq_name: gb_est18:AV403126
```

```
seq_documentation_block: 650 bp mRNA EST 06-FEB-2000
LOCUS AV403126 Bombyx mori middle silk gland 5th-instar larva Bombyx mori
DEFINITION CNA clone msgV0072 T3, mRNA sequence.
ACCESSION AV403126
VERSION AV403126.1 GI:6907214
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Plecoptera; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
Subgroup: Bombycoidea; Bombyx.
1 to 650
ORIGIN
1 cna
2 (2000)
```

Group of Radiological Sciences
Age, Chiba 263-8555, Japan
no. jp
nal, sequence direction:sequenced from T3 primer
see 'Silkbase',
ac.jp/silkbase/> for whole ESTdb.
ers
1*
mori"

```
/clone_11b="Bombyx mori middle silk gland 5th-instar larva"
/sex="female/male mixed"
/tissue_type="middle silk gland"
/dev_stage="5th-instar larva"
/note="donated by Dr. Sernal, Czech"
BASE COUNT 187 a 175 c 164 g 124 t
ORIGIN
alignment_scores:
quality: 126.00 length: 192
ratio: 1.156 gaps: 7
Percent Similarity: 56.771 Percent Identity: 28.125
alignment_block:
US-09-428-122-2 x AV403126
Align seg 1/1 to: AV403126 from: 1 to: 650
109 AsnSerValAlaAspLySerThrThrPheIleGlyPheSerSerle 125
||| ||||| ||||| |||||
16 AspTCAGATACACATGATGATGATGATGATGATGATGATGATG 65
125 userPheIleAspProGlySerSerIleThrThrglyslAlaIy 142
66 CGCGGTAGCGACACTTATGATGATGATGATGATGATGATGATG 115
142 a1sercysertThrglySerleuserleuthrlyAsnValserleu 158
||| ||||| ||||| |||||
116 TATCA...TCACCGCGCAGTTCACGATACACT..... 144
159 PheserlyAsnPheserThraspanglyAlaIlethrlalysrth 175
||| ||||| ||||| |||||
145 .....GATGACAGACAGACCTTCAGATTCACATCGGCGG 185
175 rleuserleuthrlyThrmetservalaleupheserlyAsnThr 192
||| ||| ||| |||
186 TACACCACTTATGGATACAGTTCACAGCT....CGTATGAGAGAG 229
192 erserlyslgylAlaIleGlnThrsaspAlaIlethrllethr 208
||| ||||| ||||| |||||
230 TATCATCCACGCGCAGTTCACATGATGACAGACAGACCTTACA 279
209 GlyAsnGln.....GlyGluValserPheserAsnThrSers 222
||| ||||| ||||| |||||
280 GGATCCAGTACATCCGCGGTACACACTATGATATAGTCACAGCA 329
222 rAspserGlyAlaIlePheThrGluAlaSerValThrIleSers 239
||| ||||| ||||| |||||
330 TCGTATGAGAGTATGTGCGCAGCTGCGAGT.....TCCAGTA 367
239 snAlaIlyValserPheIleAspAsnlyValThrglyAlaSer 255
||| ||||| ||||| |||||
368 ACACGTATGACAGACACAGAGATCCACACGTCGCTGATACAGC 417
256 ThrThrglyAspMetserGlyAlaIleCysAlaIlyThrSert 272
||| ||| ||| |||
418 ACTGAGGATATAGTTCC.....AGTAGCA 443
272 rAspThrlyValThrleuthrGlyAsnGlnMetleuPheSers 289
||| ||||| ||||| |||||
444 TGATGAGAGCTTACA.....TCCACCG 466
289 snrThserThrAlaGlyAla 297
||| ||||| ||||| |||||
467 ACGGTCCAGCACAGTGGAGAGCT 492
seq_name: gb_est9:AI166701
seq_documentation_block: 693 bp mRNA EST 03-DEC-1998
LOCUS AI166701
DEFINITION xylem_est.507 Poplar xylem Lambda ZAP11 library Populus balsamifera
subsp. trichocarpa cDNA 5', mRNA sequence
```

```

01.1 GI:3857985
Populus balsamifera subsp. trichocarpa.
Populus balsamifera subsp. trichocarpa.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Malviales; Saliaceae; Populus.
1 (bases 1 to 693)
REFERENCE
  Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg,
  A., Amin, B., Bhalerao, R., Larsson, M., Villarreal, R., Van Montagu,
  M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W., Gustafsson, P.,
  Uhlen, M., Sundberg, B. and Lundberg, J.
  Gene discovery in the wood-forming tissues of poplar: Analysis of
  5,692 expressed sequence tags
  Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
JOURNAL
  MEDLINE
  CONTACT: Boerjan W
  Department of Genetics
  Flanders Interuniversity Institute for Biotechnology, University of
  Gent
  Coupesackerstraat 35, B-9000 Gent, Belgium
  Tel.: +32 9 2645202
  Fax: +32 9 2645346
  Email: wboerjan@ugent.ac.be
  Seq. primer: F3 primer
  High quality sequence stop: 693.
FEATURES
  SOURCE
    1..693
    /organism="Populus balsamifera subsp. trichocarpa"
    /cultivar="trichobol"
    /db_xref="taxon:3694"
    /clone_lib="Poplar xylem Lambda ZAPIT library"
    /tissue_type="young, developing xylem"
    /dev_stage="3-year-old actively growing tree (harvested in
    May)"
    /lab_host="E.coli"
    /note="Vector: pBluescript SK-; Site1: EcoRI (5' end cDNA
    ); Site2: XhoI (3' end cDNA); Young developing xylem was
    harvested from 3-year-old, actively growing Populus
    trichocarpa 'trichobol' by peeling off the bark first and
    then scraping on the remaining xylem. cDNA was prepared
    and cloned directionally into Lambda ZAPIT. Plasmid clones
    of individual Lambda clones were obtained by in vitro
    excision."
  BASE COUNT
    221 a 142 c 192 g 138 t
  ORIGIN
    alignment_scores:
      Quality: 125.00 Length: 206
      Ratio: 1.008 Gaps: 6
      Percent Similarity: 60.194 Percent Identity: 27.184
  alignment_block:
    US-09-428-122-2 x A1166701/rev ...
  Align seg 1/1 to reverse of: A1166701 from: 1 to: 693
70 ThTAlaIleThrLysSerCysPheAsnAsnThrLysGlyAspLeuThrPh 86
|||||
583 ACTGGGCTTCCTGTTCAACCTCTCTCTCTGTTTAAAGGCTTCTTTC 534
|||||
86 eThrGlyAsnGlyAsnSerLeuLeuPheGlnThrValAspLacGlyThv 103
|||||
533 TTCCTTACTCTCTCTCTCTGCGCTTGACCTCAACCTCCGCTTCTT 484
|||||
103 AlaAlaGlyAlaAlaValAsnSerSerValValAspLysSerThrThrPhe 119
|||||
483 CTGGCGTTCCTCCACGAGCTCTGTAGTCTCTTACCTCTCTCAACA... 437
|||||
120 IlegLysSerSerLeuSerPheIleAlaSerProGlySerSerIlePh 136
|||||
01.1 GI:3857985
436 .....ACAGGCTCTACGGGTCAC 420
136 rThrGlyLysGlyAlaValSerCysSerThrGlySerLeuSerLeuThrL 153
|||||
419 TTCGGGCTCTCTTCTTCTCTCTCAACACCGGCTCT ..... 383
153 yAsnValSerLeuLeuPheSerLysAsnPheSerThrAspAsnGlyGly 169
|||||
382 ..GGGTTTCCTTAGCTGTCTCTCTCAAGTTTCTCTACTACTGTGGC 335
|||||
170 AAlaIleThrAlaLysThrLeuSerLeuThrGlyThrMetSerAlaIe 186
|||||
334 TCTCTGACGAGTCTGTGCTCT .....TCGAGCACCTCTT 297
186 uPheSerGlyAsnThrSerSerLysLysGlyAlaAlaIleGlnThrSera 203
|||||
296 ACTTCACGCTTCAGCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 247
203 sPAlaLeuThrIleThrGlyAsnGlnGlyAlaSerPheSerAspAsn 219
|||||
246 ATGCGTCACTTATCT .....GCTCTTTGGCTCTTCA 212
220 ThrSerAspSerGlyAlaAlaIlePheThrGlnAlaSerValThrI 236
|||||
211 TGAGTACTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 165
236 eSerAsnAsnAlaLysValSerPheIleAspAsnLysValThrGlyAla 253
|||||
164 CTCACCTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 118
253 eSerSerThrThrGlyAspMetSerGlyAlaAlaIleCysAlaThrLys 269
|||||
117 GCGGTGACACACCTCAACAGTGCCTTGAACAGAGAGAAATAAAG 68
270 ThrSerThrAspThrLys 275
|||||
67 CACACGTGCGATCAAAA 50
seq_name: gb_est1.A1517569
seq_documentation_block:
  LOCUS A1517569 613 bp mRNA EST 16-MAR-1999
  DEFINITION GH28575.5prine GH Drosophila melanogaster head POT2 Drosophila
  melanogaster cDNA clone GH28575 5prine similar to L05080. Cpn
  FB990010218 PID:g1517072 SWISS-PROT:O02910, mRNA sequence.
  ACCESSION A1517569
  VERSION A1517569.1 GI:4420669
  KEYWORDS EST.
  SOURCE
    ORGANISM
      Drosophila melanogaster
      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
      Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
      1 (bases 1 to 613)
  REFERENCE
    Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein,
    P., Lewis, S. and Rubin, G.M.
    BDGP/HMI Drosophila EST Project
    Unpublished (1997)
  TITLE
    G.M. Rubin-Molecular and Cell Biology
    University of California at Berkeley
    595 LSN, Berkeley, CA 94720-3200, USA
    Fax: 510 643 9349
    Email: http://www.fruitfly.org/EST. estfruitfly.berkeley.edu
    Plate: 285 Low G column: 3
    High quality sequence stop: 476.
FEATURES
  SOURCE
    1..613
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone_lib="GH28575"
    /clone_lib="GH Drosophila melanogaster head POT2"
    /sex="male and female"

```



```

338 rleuserlaspser 343
1 .....
905 TTCTGTTCTCTCTCC 920

seq_name: gb_gss9:A0640209

seq_documentation_block: 579 bp DNA GSS 08-JUL-1999
LOCUS A0640209
DEFINITION 927Pl-2F1_TV 927Pl Trypanosoma brucei genomic clone 927Pl-2F1, DNA
sequence.
ACCESSION A0640209
VERSION A0640209.1 GI:5116919
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 579)
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ull, E., Melville, S., Donalson, J.,
Fraser, C. and Adams, M.
Determination of clone end sequences from Trypanosoma brucei TREU
927/4 Pl library
Unpublished (1999)
Other_GSSs: 927Pl_2F1.TP
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
7712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208
Email: nelsayed@tigr.org
For clone/filer availability, please contact Sara Melville
(emb50@ole.bio.cam.ac.uk). Pl end sequences search page:
http://www.tigr.org/cdb/mbd/ibdb/.
Seq primer: T7
Class: Pl ends.
Location/Qualifiers
1..579
/organism="Trypanosoma brucei"
/strain="TREU927/4"
/db_xref="taxon:5691"
/clone="927Pl-2F1"
/clone_1bp="927Pl"
/note="Vector: pAD10sacBII; Site 1: Bam HI. Constructed by
Sara Melville, University of Cambridge, UK and Nancy
Shepherd, Dupont Merck, Wilmington, DE. Genomic DNA was
isolated from Trypanosoma brucei (stock TREU927/4) and
partially digested with Sau 3AI. DNA fragments were cloned
into the Bam HI site of pAD10sacBII vector (Genbank
accession U09128). The average insert size is 65 kb.
Coverage: approx 4.4 x the haploid non-mitochondosomal
genome"
BASE COUNT 114 a 181 c 101 g 183 t
ORIGIN
alignment_scores:
Quality: 123.00 Length: 170
Ratio: 1.025 Gaps: 4
Percent Similarity: 70.588 Percent Identity: 27.059
alignment_block:
US-09-428-122-2 x A0640209 ..
Align seg 1/1 to: A0640209 from: 1 to: 579
95 pNegIntrValaspAlaGlyThrValAlaGlyAlaValaIaaspser 111
|||||
77 TTCAATTCAGTTAGTACACATCTCTGCTCAAGTGAAGCTATATGC 126
|||||
111 rValValaspIysSerThrThrPheIleGlyPheSer..... 124
|||||

```

FEATURES

source

Location/Qualifiers

1..579

/organism="Trypanosoma brucei"

/strain="TREU927/4"

/db_xref="taxon:5691"

/clone="927Pl-2F1"

/clone_1bp="927Pl"

/note="Vector: pAD10sacBII; Site 1: Bam HI. Constructed by

Sara Melville, University of Cambridge, UK and Nancy

Shepherd, Dupont Merck, Wilmington, DE. Genomic DNA was

isolated from Trypanosoma brucei (stock TREU927/4) and

partially digested with Sau 3AI. DNA fragments were cloned

into the Bam HI site of pAD10sacBII vector (Genbank

accession U09128). The average insert size is 65 kb.

Coverage: approx 4.4 x the haploid non-mitochondosomal

genome"

BASE COUNT 114 a 181 c 101 g 183 t

ORIGIN

alignment_scores:

Quality: 123.00 Length: 170

Ratio: 1.025 Gaps: 4

Percent Similarity: 70.588 Percent Identity: 27.059

alignment_block:

US-09-428-122-2 x A0640209 ..

Align seg 1/1 to: A0640209 from: 1 to: 579

95 pNegIntrValaspAlaGlyThrValAlaGlyAlaValaIaaspser 111

|||||

77 TTCAATTCAGTTAGTACACATCTCTGCTCAAGTGAAGCTATATGC 126

|||||

111 rValValaspIysSerThrThrPheIleGlyPheSer..... 124

|||||

```

127 AGTTCTTTTCAGATGATATAAGGCTGATTTTCTCTAATGCTTA 176
125 ..LeuSerPheIleAlaSerProGlySerSerIleThrThrGlyGly 140
177 CCTTACCTGACGATGCTGATGCTGAGCTGACCTGACGCGCGG 226
141 AlaValSerCysSerThrGlySerLeuSer.....LeuThrLy 153
227 CTATATACCTCCGCGCACACACTCTCTCGACAGCCTTGCTCACTC 276
153 SAsnValSerLeuPheSerIleGlySAsnPheSerThrAspAsnGly 170
277 CTGCGGAGCTCTATCTCTCGGACCTTCAGACCTTCGCTTCGCA 326
170 IalIeThrAlaLysThrLeuSerLeuThrGlyThrMetSerAlaLeu 186
327 GTTTCATTTCTGGGAGCTTCAAGCTTCTCGACAGTTCTCA 376
187 PheSerIleAlaSerThrSerSerIleGlyGlyAlaIleGlnThrSer 203
377 TTCTGGGAGCTTCAAGCTTCTCGGACCTTCATCTCATCTCGG 426
203 PAlaLeuThrIleThrGlySAsnGlyValSerPheSerAspAsn 220
427 AGCTTCTAGC...TCAGCTTCTCGACAGTTCTCATCTCGGACACT 473
220 hrSerSerAspSerGlyAlaIlePheThrGlyAlaSerValThrIle 236
474 TCAGCTCAGCTCTCTCGACAGTTCTCATCTCGGAGCTTCAGCTCA 523
237 SerAsnAlaLysValSerPheIleAspAsnLysValThrGlyAla 253
524 AGCTCTCTCGACAGTTCTCATCTC.....TCGGACACTTCAGCTC 564
253 rSerSerThr 256
565 AAGCTCTCT 574
seq_name: gb-est8:A115059

```

```

seq_documentation_block:
LOCUS      A115059      820 bp      mRNA      EST      02-SEP-1998
DEFINITION ui41h04.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:1885015.5' similar to gb:J04634.crn3 Mouse cell surface
antigen 114/A10 mRNA, complete (MOUSE);, mRNA sequence.
ACCESSION  A115059
VERSION    A115059.1 GI:3515383
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM  Mus musculus.
REFERENCE  1.
AUTHORS   Mammali, J., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
          Lucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
          K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
          Wylie, T., Lennon, G., Soares, B., Wilson, R. and
          Mouse EST Project

```

Mammali, J., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Lucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Mouse EST Project
 Mammali, J. et al.
 School of Medicine
 Box 8501, St. Louis, MO 63108
 royalti-free through LNL; contact the
 (age.lnl.gov) for further information.
 133.

```

source
1. .820
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1885015"
/clone_lib="Sugano mouse embryo mewa"
/seq_stage="embryo, 14 dpc"
/lab_host="DH10B"
/note="Vector: pME18-FL3; site 1: DraIII (CACTGCTG);
site 2: DraIII (CACCATGCT); 1st strand cDNA was primed
with an oligo(dT) primer (ATGCGCCCTTTTCTTTTCTTTT);
double-stranded cDNA was ligated to a DraIII adaptor
(TGTTGGCCCTGCTG), digested and cloned into distinct DraIII
sites of the pME18-FL3 vector (5' site CACTGCTG, 3' site
CACCATGCT). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTCTGCTCTAAGCTGCG and 3' end primer
CGACTGCGCTCGAGCACA."
BASE COUNT      172 a      269 c      167 g      212 t
ORIGIN

```

```

alignment_scores:
Quality: 121.00      Length: 247
Ratio: 58.400      Gaps: 11
Percent Similarity: 58.300      Percent Identity: 27.126

```

```

alignment_block:
US-09-428-122-2 x A115059

```

```

Align seg 1/1 to: A115059 from: 1 to: 820

```

```

115 lysSerThrThrPheIleGlyPhe.....SerSerLeuSerPheIle.. 128
26 AGAGCAGCTTCATGATGAAGCTTCGCTCTCTCTCATGCTCTCTTT 75
129 .....AlaSerProGlySerSerIleThrThrGlyGlyG 140
76 GGTAGCGGTGGAGTGCATCTCGAGCTCTGATGACATGCTCTCTG 125
140 lYAlaValSerCysSerThrGlySerLeuSerLeuThrLysAsnValSer 156
126 GTGGCACCAGCTCTCTCCACAGCATCAAGTCAATCTCCAGTAGTTCA 175
157 LeuLeuPheSerIleGlySAsnPheSerThrAspAsnGlyAlaIleThr 173
176 TCTAGGCTCTACTACATGCTCT...TCTGTGGCCGAGCTCTCC 222
173 aLysThrLeuSerLeuThrGlyThrThrMetSerAlaLeuPheSerGly 190
223 CACACCGTACAGACTCATCTCCAGTAGTTATCTCAGGCTCTACTA 272
190 snThrSerSerIleGlyAlaIleGlnThr..... 201
273 CACATGCTCTCTGCTGGCGCACGCTCTCCACACGAGTACAGTCAA 322
202 .....SerAspAlaLeuThrIleThrGlySAsnGlyG 213
323 TCTCCAGTAGTTCACTTCAGGCTCTCTACACATGCTCTCTGCTGG 372
213 uValSerPheSerAspAsnThrSerSerAspSerGlyAlaIlePhe 230
373 CGCCAGTCTCTCCACACGAGTACAGTCAATCTCCAGTAGT...TCAT 419
230 hrGlyAlaSerValThrIleSerAsnAlaLysValSer..... 243
420 CTCAGGCTCTTACTACATGCTCTCTGCTGGCGCACCTCTCCACC 469
244 PheIleAspAsnLysValThrGlyAlaSerSer.....SerThrThrG 258

```

```

519 ..CAATCTCCAGGAGTATCTAGGCTCTACTACAC
520 ..SerGlyValAlaIleCysAlaTyrIlePheSerThrSphTrl 275
521 ..|||||
522 ..AGTCTCCACACAGGTACAGAGTCC 563
523 ..ATCGCTCTCTGTGTGGCCG
275 ysValThrIleuThrGlyAsnGlnMetLeuPheSerAsnThrSer 291
564 AATCTCCAGGAGTATCTAGGCTC.....TCTACTACACA 601
292 ThrThAlaGlyGlyAlaIleTyrValLysLeuGluLeuAlaSerG1 308
602 TCTCTCTGTGTGGCCGAGTCTCCACCACGCTACAGAGTCACTCC 651
308 yGlyLeuThrIleuPheSerArg...AsnSerValAsnGlyLys.....321
652 AGGTACTTCTCTAGGCTCCAGCCCACTCACTCTGTGTGTGGCCAGCT 701
322 .....ThrAlaProLysGlyAlaIleAlaIleGluAsp 333
702 CTTCACCGCTGCTAGCGGTGAGACAGAGTTCACAGTAT 742

seq.name: gb_gss23:CNS03TF7

seq documentation block: 918 bp DNA 17-May-2000
LOCUS CNS03TF7
DEFINITION Tetradon nigroviridis genome survey sequence PUC-Orl end of clone
056109 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL259756.1 GI:7980768
VERSION AL259756.1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Euplaterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
Tetraodontiformes; Tetraodontidae; Tetraodon.
1 (bases 1 to 918)
REFERENCE 1 Roest-Crolius,H., Jallion,O., Dasilva,C., Fizes,C., Fisher,C.,
AUTHORS Bouneau,B., Billault,A., Queller,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 918)
AUTHORS Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,B., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brothier,F., Queller,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 918)
AUTHORS Genoscope.
TITLE Direct Submission
COMMENT Submitted (12-Apr-2000) to the EMBL/Genbank/DBAP databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetradon.
FEATURES
Source
1. 918
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="056109"
/clone_lib="G"
/note="Genoscope sequence ID : C0B9056CF05SP1-end :
PUC-Orl"
BASE COUNT 293 a 324 c 136 g 159 t 6 others

```

```

alignment_scores:
  quality: 121.00      length: 301
  ratio: 0.752        gaps: 12
  percent similarity: 53.488  percent identity: 23.588
alignment block:
US-09-428-122-2 x CNS03TF7
Align seq 1/1 to: CNS03TF7 from: 1 to: 918
100 AlAGyThrValAlaGlyAlaValAsnSerSerValAlaSpLyse 116
|||||
72 GCAGCTCTCTGTACACCATCAGCAGCAACACTGCGATACAGCTAC 121
116 rThrThrPheIleGlyPheSerSerLeuSerPheIleAlaSerProGlyS 133
122 AACTCA.....GAACTGTGGAG 138
133 eSerIleThrThrGlyLysGlyAlaValSerCysSerThrGlySerLeu 149
139 GCACCAACAACCTCTGTCTCCACACAGACAGCAACACACTGATGTA 188
150 SerLeuThrLysAsnValSerLeuPheSerLysAsnPheSerThrS 166
189 GCACACTACAGCTGAGCTACACAGCTGCCACTCCACT...TCCATATCA 235
166 pAsnGlyGlyAlaIleThrAlaLysThrIleuSerLeuThrGlyThrThm 183
236 GACTTATCAGCAGAACTTATCCACCATCAACCCACCAACCAAAACCA 285
183 eSerIleAlaPheSerGluAsnThrSerSerLysGlyGlyAlaIle 199
286 CACACAGACGCTCACAACCAACATCATCA.....GGTACCTC 326
200 GlnThrSerAspAlaLeuThrIleThrGlyAsnGlnGlyValSerP 216
327 ACTACATATGAH.....ACAGCAACAACCTTTCAG.....ACCTT 361
216 eSerAspAsnThrSerSerAspSerGlyAlaIlePheThrGluAs 233
362 CTACCATCATCATCAGCATCAAGT..... 389
233 eValThrIleSerAsnAsnAlaLysValSerPheIleAspAsnLysVal 249
390 ..GTAAACAATGTAACACAYMCTCCACTCCCTCTGCTCCACCTG 437
250 ThrGlyAlaSerSerSer.....ThrTh 257
438 ACTCCAAACCAAGTGCAGCTCCAGCTTAACACACACACACATGAGAGTAC 487
257 rGlyAspMetSerGlyGlyAlaIleCysAlaTyrIlePheSerThrAspT 274
488 AATCCCAATCATCATCAGTGAAGTCACTACATGGAAGGCAACAACTT 537
274 hrLysValThrIleuThrGlyAsn.....GlnMetLeuPheSer 287
538 CTGAGAGCGCTTTAAACATCGACTCCACAACACAGAGGTACAACTAG 587
288 AsnAsnThrSerThr.....ThrAlaGlyGlyAla 297
588 AAGGAGCCTCTCACTCACTCTACTCTCCACCTGACTCCAAACAGTGC 637
297 AleIleThrLysLysLeuGluLeuAlaSerGlyLysLeuThrPheS 314
638 AGCTCTGTGTACCA...TCTTACGACACACTGAACTTCACTCTCCG 684
314 eArgAsnSerValAsnGlyLysThrAlaProLysGlyGlyAlaIleAla 330
685 CACTTCACAGACAGCAAGGAGGACAGAGTACGCGGTTCATCTGCA 734
331 IleGluAspSerGlyLysLeuSerSerAlaAspSerGlyAspIleVa 347
735 GTAGCAGCAACCAAGCAGCAACCAAGAACTTCTGTTCATCAGACACCA 784

```

347 lPheLeuGlyAsnThrValThrSerThrThrProGlyThrAsnArgSer. 363
: : : : :
785 C.....AGCAGACTGTGTAGAAAGCACAAGCTGGAGRACCACTGCAGCC 828
364SerIleAspLeuGlyThrSerAlaLysMethThr 374
829 CCAGCTCAGACACTCCACACAGGGGTATATACTTCAGCCTCCACAGCC 878
375 Ala 375
111
879 GCT 881

470 ACGGTACAGAG
258 YASPK
520
Fri Nov 24 13:50:17 2000